Volume 45 January 1959 Number 1

Proceedings of the
National Academy of Sciences
of the United States of America

The Proceedings of the National Academy of Sciences

OFFICERS OF THE ACADEMY

DETLEV W. BRONK
President

FARRINGTON DANIELS
Vice President

H. P. ROBERTSON
Foreign Secretary

HUGH L. DRYDEN
Home Secretary

WILLIAM J. ROBBINS
Treasurer

EDITORIAL BOARD OF THE PROCEEDINGS

WENDELL M. STANLEY
Chairman

HUGH L. DRYDEN, *Home Secretary*
WILLIAM J. ROBBINS, *Treasurer*

H. P. ROBERTSON, *Foreign Secretary*
EDWIN B. WILSON, *Managing Editor*

GREGORY BREIT	THOMAS FRANCIS, JR.	H. W. MAGOUN
HARRISON BROWN	C. H. GRAHAM	HARLOW SHAPLEY
ERNST CLOOS	SAUNDERS MAC LANE	PAUL WEISS

Subscription price is $12.50 for 1 year, $24.00 for 2 years, and $34.50 for 3 years; the price of a single issue is $1.50. Make all remittances payable to the National Academy of Sciences in U.S. currency or its equivalent.

Editorial correspondence should be addressed to the PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES, 2101 Constitution Avenue, Washington 25, D. C.

Business correspondence should be addressed to the Printing and Publishing Office, National Academy of Sciences, 2101 Constitution Ave., Washington 25, D. C.

Subscribers are requested to notify the Printing and Publishing Office of the Academy and their local postmaster immediately of change of address. Notices to the Academy should provide both the old and the new address.

Microfilms of complete volumes of this journal are available to regular subscribers only and may be obtained at the end of the volume year from University Microfilms, 313 N. First Street, Ann Arbor, Michigan.

Second-class postage paid at Easton, Pennsylvania.

PRINTED IN THE U. S. A.

THE PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES
is published monthly by
THE NATIONAL ACADEMY OF SCIENCES

Proceedings of the
NATIONAL ACADEMY OF SCIENCES

Volume 45 · Number 1 · January 15, 1958

*A MITOCHONDRIAL FACTOR PRODUCING "LOOSE-COUPLING" OF RESPIRATION**

By LeMar F. Remmert† and Albert L. Lehninger

DEPARTMENT OF PHYSIOLOGICAL CHEMISTRY, JOHNS HOPKINS SCHOOL OF MEDICINE, BALTIMORE

Communicated November 10, 1958

Carefully prepared liver or kidney mitochondria yield high rates of respiration and maximum P:O ratios in the presence of adenosine diphosphate (ADP) and phosphate but respire only very slowly in the absence of ADP, as was first shown by Lardy and Wellman.[1] Such "tight-coupling" of respiration has since been observed in many laboratories and has led to the concept[1,2] that the ADP level is an important determinant of the rate of respiration. On the other hand, some types of mitochondrial preparations, although capable of yielding high P:O ratios in the presence of ADP, do not require the presence of ADP for high rates of respiration. An example is provided by the fly sarcosomes described by Sacktor and Cochran.[3] Phosphorylating respiration of the latter type may be designated as "loosely coupled." There is very little information available concerning the factors responsible for "tightness" or "looseness" of respiratory coupling, the mechanism of these effects, or their physiological significance.

In this communication we wish to report that a soluble protein-containing fraction obtained from rat liver mitochrondria is capable of fully releasing the respiration of mitochondrial subfragments from its dependence on the presence of ADP, without lowering the P:O ratio greatly when ADP is available. This mitochondrial factor, designated here as "R factor" for its respiration-releasing activity, is thus capable of converting a "tightly coupled" phosphorylating respiration to a "loosely coupled" system.

The mitochondrial fractions containing the R-factor activity were prepared as follows: Freshly isolated rat liver mitochondria, following washing with 0.25 *M* sucrose, were suspended in 0.05 *M* sucrose–0.02 *M* KCl and subjected to sonic disruption at 9 kilocycles at 0° for 60 minutes. The sonicate was adjusted to pH 7.4, clarified by sedimentation at 115,000 × *g* for 30 minutes, dialyzed against several changes of 0.03 *M* KCl for 40 hours, again adjusted to pH 7.4, and clarified with a high-speed centrifugation. This dialyzed extract, or a protein fraction precipitating from it at pH 5.5, was used in the experiments described. The action of the R factor was assayed on membrane fragments of rat liver mitochondria capable of phosphorylating electron transport;[4–6] under the conditions used in these experiments the oxidation of β-hydroxybutyrate (BOH) by oxygen is ordinarily stimulated between 60 and 400 per cent by the presence of an excess of ADP.[7]

Figure 1 shows that addition of increments of the R-factor preparation to a suspension of mitochondrial fragments oxidizing BOH in the absence of ADP produces increasing rates of respiration, until a plateau is reached, at which further addition of R factor produces no further increase in rate of oxidation of BOH. This plateau coincides closely with the maximum rate of respiration which is given by the presence of an excess of ADP as phosphate acceptor; it coincides also with the maximum rate of respiration given by the presence of concentrations of 2,4-dinitrophenol (DNP) which uncouple phosphorylation completely. R-factor preparations thus are capable of full release of respiration from dependence on ADP when added to the submitochondrial preparations. The R-factor preparations, when tested in the system alone in the absence of the mitochondrial subfragments, show no ability to oxidize BOH.

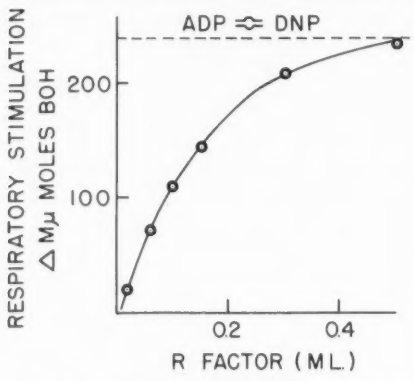

FIG. 1.—The test system contained 0.015 *M* phosphate, pH 6.5, 0.01 *M* DL-BOH, and mitochondrial fragments (286 μg of N) in a total volume of 2.0 ml., incubated 17 minutes at 22°. The R-factor preparation contained 2.56 mg. of N per ml. The data are corrected for the amount of BOH oxidized by the mitochondrial fragments alone; respiratory release by R factor is compared with that given by ADP (0.0024 *M*) or DNP (1×10^{-5} *M*).

The stimulation of respiration caused by R factor is qualitatively different from that caused by DNP. The data summarized in Table 1 compare the action of

TABLE 1

EFFECTS OF R FACTOR AND DNP ON RESPIRATION AND PHOSPHORYLATION*

	Addition	ADP	Respiration (mμmoles BOH Oxidized)	P:O
1	—	−	135	..
	—	+	223	1.45
2	R factor	−	240	..
	R factor	+	263	1.24
3	DNP, 3×10^{-5}M	−	240	..
	DNP, 3×10^{-5}M	+	237	0.02

* The test system contained 0.015 *M* phosphate (7.2×10^8 c.p.m. of P^{32}), pH 6.5, 0.01 *M* DL-BOH, and mitochondrial fragments (220 μg. of N) in a total volume of 2.0 ml. Where added, the ADP concentration was 0.0024 M; the R factor contained 384 μg. of N. The reaction period was 15 minutes at 22°.

such concentrations of the R factor and of DNP as just produce full release of respiration, tested in the presence and absence of ADP. It is seen that both R factor and DNP stimulate respiration of the ADP-less system to the level given by ADP alone. However, in the presence of excess ADP, the R factor causes very little uncoupling of phosphorylation, whereas DNP uncouples nearly completely. It is therefore evident that the addition of the R factor has converted the relatively "tightly coupled" system of Experiment 1, which requires ADP for a maximum rate of respiration, into the "loosely coupled" phosphorylating system of Experiment 2, which has no dependence on ADP for respiration. The P:O ratios in the "tightly coupled" and "loosely coupled" systems are approximately the same.

On the other hand, DNP releases respiration but produces an "uncoupled" system.

The respiration released by the presence of the R factor is completely inhibited by antimycin A in low concentrations, as is the respiration in the presence of ADP.[4] Tests of the site of action of the factor in the respiratory chain indicate that it stimulates electron transfer not only in the span BOH to cytochrome c[8] but also in the span cytochrome c to oxygen, tested with ascorbate as non-enzymatic reductant of cytochrome c[9]. It therefore appears likely that the factor can act at all three coupling sites in the respiratory chain. The R factor also stimulates the oxidation of BOH by freshly prepared *intact* rat liver mitochondria in the absence of ADP. On the other hand, the respiration of non-phosphorylating mitochondrial fragments similar to Keilin-Hartree preparations is not stimulated by the R factor. The ability to release respiration in the submitochondrial fragments in the absence of ADP, without severe uncoupling when ADP is present, is relatively unique; it is not shown by DNP, Dicumarol, gramicidin, ethylenediaminetetraacetate, α-tocopherol, vitamin K_1, menadione, ribopolynucleotides, serum albumin, coenzyme A, or thyroxine; nor can antimycin A–sensitive respiration be released by the addition of known coenzymes such as diphosphopyridine nucleotide. The action of the R factor is not prevented by addition to the test system of Mg^{++}, the lack of which can cause loss of respiratory control and decline in P:O ratios in intact mitochondria.[10]

The characteristic activity of the R factor in increasing respiration is not occasioned by the simultaneous presence of ADP and an intense ATP-ase activity in the preparations. The exhaustively dialyzed material contains no ADP or ATP detectable enzymatically or chromatographically. After precipitation at pH 5.5 the R-factor preparations have no ATP-ase activity and stimulate the ATP-ase of the mitochondrial fragments only moderately. The data in Table 2 demonstrate

TABLE 2

NON-INVOLVEMENT OF ADP IN RESPIRATORY STIMULATION BY R. FACTOR*

Addition	Hexokinase and Glucose	Respiration (mμmoles BOH Oxidized)	Phosphorylation (mμmoles ATP^{32})	P:O
.....	−	202	0	0.0
.....	+	199	27	0.14
R factor, 0.4 ml.	−	403	0	0.0
R factor, 0.4 ml.	+	422	3	0.01
ADP, 8×10^{-6} *M*	−	289	36	0.12
ADP, 8×10^{-6} *M*	+	404	634	1.57
ADP, 2.4×10^{-3} *M*	−	445	704	1.59
ADP, 2.4×10^{-3} *M*	+	455	838	1.84

* The test system contained 0.015 *M* phosphate (1.03×10^6 c.p.m. of P^{32}), pH 6.5, 0.01 *M* DL-BOH, and mitochondrial fragments (262γ of N) in a total volume of 2.0 ml. Where added, the R factor contained 1056 γ of μg.; the hexokinase concentration was 0.5 mg. (270 K.M. units[18]) per ml.; the glucose concentration was 0.01 *M*.; Incubated 15 minutes at 22°.

that very large excesses of highly purified yeast hexokinase and glucose added to membrane fragments whose respiration has been stimulated by an excess of the R factor in the absence of added ADP causes no trapping of esterified phosphate. However, the amount of hexokinase added is capable of trapping phosphate efficiently at levels of ADP as low as 8×10^{-6} *M*. It is evident that the release of respiration by the R factor is not caused by catalytic action of traces of ADP coupled with intense ATP-ase activity; presumably the characteristic action of the R factor occurs at some stage in the coupling mechanism prior to formation of ATP.

The R-factor preparations contain a mixture of factors which affect the coupling of respiration with phosphorylation, since uncoupling of phosphorylation can also be produced when great excesses of the preparation are added to the standard test system, far beyond the amoung required to release respiration maximally. Furthermore such great excesses of the R factor also stimulate ATP-ase activity[11] and inhibit the ATP-P_i^{32} exchange reaction.[12] This fact suggested that the R-factor preparations may contain the isoöctane-soluble uncoupling factor present in mitochrome and aged mitochondrial preparations which Hülsmann *et al.* have recently described.[13] Isoöctane extracts of the R-factor preparations were found to stimulate respiration in the test system described. However, such extracts also caused substantially complete uncoupling of phosphorylation when ADP was present; these effects were found to be simulated by oleic acid.[14] The isoöctane-extractable factor alone is therefore not responsible for the production of "loosely coupled" phosphorylating respiration. Although participation of the isoöctane-soluble material in producing "loose" coupling of respiration as here described is not excluded, other experiments indicate that a protein fraction is necessary. Separation and purification of the factors involved is under way.

Freshly prepared rat liver mitochondria contain very large amounts of R-factor activity. For example, the amount of activity extracted by sonication from fresh mitochondria can produce complete loss of respiratory control in an amount of membrane fragments derived from about three times the equivalent amount of mitochondria. However, since freshly prepared mitochondria are "tightly" coupled, the R factor must be present in a latent or precursor form.

These experiments define three extreme poles of behavior of respiratory systems. In "tightly coupled" respiration, as in freshly prepared normal mitochondria, there is very little respiration in the absence of ADP, and the P:O ratio is maximum when ADP is present. In "loosely coupled" respiration the P:O ratio may be maximal or nearly maximal in the presence of a phosphate acceptor system, but the rate of respiration is as high in the absence of ADP as in its presence. The third extreme case is represented by fully "uncoupled" respirations, such as produced by DNP, in which ADP is not necessary for maximum respiration, nor is there any phosphorylation when ADP is present. Intermediate degrees of behavior are, of course, possible. It is important to note that both "loose' coupling and "uncoupling" can be produced in mitochondrial fragments by adding substances derived from mitochondria, namely the R-factor preparations described here and the isooctane-extractable factors described by Hülsmann *et al.*,[13] respectively. These mitochondrial factors may thus be involved in intracellular or intramitochondrial control of respiration and phosphorylation.

It is of significance that Hoch and Lipmann have found[15] that mitochondria isolated from the livers of thyrotoxic animals have a normal P:O ratio when tested in the presence of a phosphate acceptor system. However, such mitochondria differ from those isolated from euthyroid animals in that they show a high and nearly maximum rate of respiration when ADP is omitted, a situation which corresponds to the "loose-coupling" produced by addition of the mitrochondrial R factor described in this paper. It is therefore suggested that the "unmasking" or release of its mitochondrial R factor may be subject to physiological or endocrine control and that it can in turn determine the response of the respiratory chain to

the ambient ADP concentration in the cell. These relationships are being examined further.

The following general reaction sequence, in which X designates a substance capable of forming a "high-energy" complex with a specific respiratory carrier during electron transport, has been postulated for the coupling mechanism: [16,17]

$$(1)\ \text{Carrier} + X \xrightarrow{\text{electron transport}} \text{Carrier} \sim X$$

$$(2)\ \text{Carrier} \sim X + P_i \rightleftharpoons \text{Carrier} + P \sim X$$

$$(3)\ \text{P} \sim X + \text{ADP} \rightleftharpoons \text{ATP} + \text{X}$$

It is generally assumed that uncoupling agents such as DNP act by causing the decomposition of Carrier $\sim X$ or an equivalent high-energy intermediate early in the coupling sequence.[2,16] Such an action could account for respiratory stimulation by DNP in the absence of ADP, as well as the uncoupling by DNP. On the other hand, the R factor stimulates respiration in the absence of ADP, without producing uncoupling when ADP is present. It may therefore be postulated that the R factor and ADP compete for the same intermediate in the coupling mechanism, such as P $\sim X$. The mechanism of action of the R factor is being investigated further by examination of its effect on the partial reactions of the coupling process.

Summary.—A protein-containing factor (R factor), which is apparently "latent" in fresh rat liver mitochondria, is released in active soluble form by sonic oscillation. It was found capable of fully releasing respiration of mitochondrial subfragments from its dependence on ADP as phosphate acceptor, without lowering appreciably the P:O ratio when ADP is also added. The R factor thus converts "tightly coupled" into "loosely coupled" phosphorylating respiration and is therefore qualitatively different in action from 2,4-dinitrophenol, which produces "uncoupled" respiration. The significance of these three poles of respiratory behavior, as well as the mode of action of the R factor, is discussed briefly.

* Supported by grants from the U. S. Public Health Service, the National Science Foundation, the Nutrition Foundation, Inc., and the Whitehall Foundation.

* U. S. Public Health Service Research Fellow of the National Cancer Institute. On sabbatical leave of absence from the Department of Agricultural Chemistry, Oregon State College, Corvallis.

[1] H. A. Lardy and H. Wellman, *J. Biol. Chem.*, **195,** 215, 1952.

[2] B. Chance and G. R. Williams, *Adv. Enzymol.*, **17,** 65, 1956.

[3] B. Sacktor and D. G. Cochran, *Arch. Biochem. Biophys.*, **74,** 266, 1958.

[4] C. Cooper and A. L. Lehninger, *J. Biol. Chem.*, **219,** 489, 1956.

[5] T. M. Devlin and A. L. Lehninger, *J. Biol. Chem.*, **233,** 1586 (1958).

[6] A. L. Lehninger, C. L. Wadkins, C. Cooper, T. M. Devlin, and J. L. Gamble, Jr., *Science*, **128,** 450, 1958.

[7] A. L. Lehninger and L. F. Remmert. Unpublished observations.

[8] T. M. Devlin and A. L. Lehninger, *J. Biol. Chem.*, **219,** 507, 1956.

[9] C. Cooper and A. L. Lehninger, *J. Biol. Chem.*, **219,** 519, 1956.

[10] H. Baltscheffsky, *Biochim. et Biophys. Acta*, **25,** 382, 1957.

[11] C. Cooper and A. L. Lehninger, *J. Biol. Chem.*, **224,** 547, 1957.

[12] C. Cooper and A. L. Lehninger, *J. Biol. Chem.*, **224,** 561, 1957.

[13] W. C. Hülsmann, W. B. Elliott, and H. Rudney, *Biochim. et Biophys. Acta*, **27,** 663, 1958.

[14] B. C. Pressman and H. A. Lardy, *J. Biol. Chem.*, **197,** 547, 1952; *Biochim. et Biophys. Acta*, **21,** 458, 1956.

[15] F. L. Hoch and F. Lipmann, these PROCEEDINGS, **40,** 909, 1954.

[16] A. L. Lehninger, *Harvey Lectures*, Series 39, 176, 1953–54.

[17] C. L. Wadkins and A. L. Lehninger, *J. Biol. Chem.*, **233,** 1589 (1958).

[18] M. Kunitz and M. R. McDonald, *J. Gen. Physiol.*, **29,** 393, 1946.

*GLYCOGEN SYNTHESIS FROM GLUCOSE, GLUCOSE-6-PHOSPHATE, AND URIDINE DIPHOSPHATE GLUCOSE IN MUSCLE PREPARATIONS**

By Phillips W. Robbins, Robert R. Traut, and Fritz Lipmann

ROCKEFELLER INSTITUTE, NEW YORK CITY

Communicated November 7, 1958

INTRODUCTION

The function of insulin as a regulator of permeability of cell membranes is, at present, in the foreground of attention. Nevertheless, experiments such as those reported in the lucid paper by Kipnis and Cori[1] on the change of the temperature coefficient for pentose transfer into diaphragm with insulin addition indicate a possible shift from diffusion to a chemical and therefore probably enzymatic reaction. Such observations suggest that the search for a possible primary metabolic effect of insulin should not yet be fully abandoned. These considerations led us to investigate pigeon breast muscle, where Krebs had found effects of insulin on respiration maintenance[2] which had been confirmed.[3] The fact that such an effect appeared to be unique with pigeon muscle seemed, essentially, not to argue against its possibly reflecting an aspect of insulin activity.

Although with the pigeon breast muscle system no insulin effect could be obtained when measuring glycogen synthesis, we continued because we became aware, during these studies, of some seemingly unresolved problems in the field of glycogen synthesis proper. Certain problematic aspects of the relation of phosphorylase to glycogen synthesis had been brought out already through work on the epinephrine effect on phosphorylase activity. This work, done mainly by Sutherland and his co-workers[4–6] and by Cori,[7] showed clearly that phosphorylase activation invariably led to glycogen breakdown and never to glycogen synthesis. Since our experiments with breast muscle homogenates gave strong support to the proposition that phosphorylase was not involved in glycogen synthesis, we considered the possibility that the synthesis might go by way of UDPG. The demonstration by Leloir[8] that liver contains an enzyme which will form glycogen from UDPG, made it seem likely that the problem of muscle glycogen synthesis could be explained in terms of a UDPG pathway. In the experiments reported below, results with the pigeon breast muscle homogenate are described, and preliminary data are given on the rabbit and rat skeletal muscle system which synthesizes glycogen from UDPG.

MATERIALS AND METHODS

C^{14}-labeled glucose was purchased from Volk Radiochemical Corporation. It was converted to glucose-6-phosphate with crystalline yeast hexokinase, the gift of Dr. M. Kunitz, and was converted to UDPG by a modification of the method of Glaser and Brown.[9] Both glucose-6-phosphate and UDPG were purified by chromatography on Dowex-1 before being used. The UDPG was free of UDP-galactose. ATP, DPN, TPN, and CoA were purchased from the Pabst Laboratories. Pigeon breast muscle kochsaft was prepared according to Krebs and Eggleston.[2]

Homogenate System for the Incorporation of Labeled Glucose into Glycogen.—Pigeons were decapitated or anesthetized with Nembutal, and a section of breast

muscle was removed rapidly and chilled. The muscle was minced and then homogenized in 4 volumes of oxygenated 0.15 *M* KCl, 0.05 *M* glyclglycine, pH 7.6. Homogenization was carried out in a glass tube with a Teflon pestle. Incubations were in Warburg flasks in O_2 at 37°, with KOH in the center well. The flasks contained 0.6 ml. of homogenate, radioactive glucose, and other additions, such as kochsaft and Na aspartate, in a final volume of 1 ml. and were incubated for 60–75 minutes. In most experiments the glucose concentration was 2×10^{-3} *M*, 100,000 cpm/ml. The reaction was stopped by removing the flask and mixing the contents with 1 ml. of 60 per cent KOH. The mixture was heated in a boiling-water bath for 30 minutes, and 0.25 ml. of saturated Na_2SO_4 and 3 ml. of 95 per cent ethanol were added. After standing on ice, the precipitated glycogen was centrifuged, redissolved in 2 ml. of water, and reprecipitated with 3 ml. of ethanol. The final precipitate was dissolved in 1 ml. of water, plated, and counted. The glycogen isolation method is very similar to the method described by Hassid and Abraham.[10] The loss due to the solubility of glycogen in the alcoholic solution was calculated to be about 50–100 μg.[11]

Glucose Incorporation into Glycogen in Pigeon Breast Muscle Homogenates.—When experiments were carried out under conditions similar to those described by Krebs[2] for the insulin effect, it was found that glucose was not incorporated into glycogen. No incorporation was observed until the system was modified by the elimination of phosphate buffer. As shown in Table 1, the incorporation depended almost completely on the addition of either a muscle kochsaft or CoA. The kochsaft-stimulated system was found to be inhibited 90 per cent by 0.01 *M* phosphate and 100 per cent by 5×10^{-5} *M* 2,4-dinitrophenol.

TABLE 1*

STIMULATION OF GLUCOSE INCORPORATION INTO GLYCOGEN BY KOCHSAFT AND CoA

Addition to Homogenate	Incorporation (c.p.m.)
None	200
Kochsaft	12,000
ATP	800
DPN and TPN	900
CoA	7,700
CoA, ATP, DPN, and TPN	6,500

* Incorporation experiments were carried out in Warburg flasks for 60 minutes in O_2, as described in the text. The amount of kochsaft added was 0.2 ml./ml of incubation. ATP concentration was 5×10^{-4} *M*, while 50 μg/ml each of DPN and TPN were added. The CoA concentration was 5 units/ml.

Since the energy supply for glucose incorporation in this system was derived mainly from oxidative phosphorylation, the effects of kochsaft and CoA became clear when it was found that at least one of these materials was required in order to maintain oxygen uptake throughout the incubation period. In later experiments, aspartic acid was also added to the incorporation system. It was found that aspartate increased the incorporation, probably by helping to maintain the adenine nucleotide concentration at a level adequate for normal oxidative phosphorylation.[12] Stimulation of incorporation by insulin was never observed. Epinephrine inhibited incorporation only at very high concentrations (5–50 μg/ml).

Effect of Phosphorylase Kinase on Glucose Incorporation System.—We now pro-

ceeded to explore the relationship of phosphorylase to this incorporation of glucose into glycogen observed in the homogenate system. Although a direct answer to this question was not likely to be obtained in the complex homogenate system, it seemed possible to get at least a partial answer by studying the effect of phosphorylase kinase. This enzyme has recently become available in a highly active state through the work of Krebs and Fischer.[13]

Muscle homogenates were supplemented with crystalline phosphorylase *b*[14] and were incubated in pairs, one with and one without phosphorylase kinase added. As can be seen in Figure 1, the addition of kinase leads to an appreciably higher steady-state phosphorylase *a* concentration than in the control. This higher phosphorylase *a* concentration, however, not only does not stimulate but, in fact, completely eliminates the incorporation of glucose into glycogen. A similar experiment is shown in Figure 2. Here labeled glucose-6-phosphate or labeled glucose-6-phosphate and kinase were added after 8 minutes of incubation. On the left, it can be seen that the kinase lowers incorporation. On the right, a corresponding increase in the rate of glycogen breakdown may be noted.

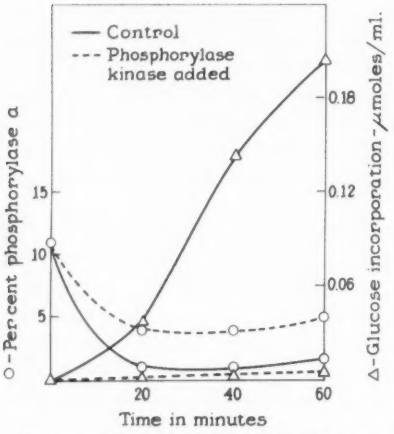

FIG. 1.—Phosphorylase kinase effect on glucose incorporation into glycogen. The incubations were carried out in Warburg flasks as described in the text with 0.2 ml. of kochsaft, 0.008 *M* Na aspartate, and 1 mg. of crystalline phosphorylase *b* (Fischer and Krebs, *J. Biol. Chem.*, **231**, 65, 1958) added to all flasks. Phosphorylase kinase was prepared according to Krebs and Fischer (*Biochim. et Biophys. Acta*, **20**, 150, 1956) and was added to a final concentration of 0.2 mg/ml. Since the enzyme is stored in a sucrose-phosphate mixture, a comparable amount of sucrose-phosphate was added to controls. The reactions were stopped by adding 0.1 ml. of 1 *M* NaF 0.1 *M* EDTA from the side arm. After small samples had been diluted for phosphorylase assay (G. T. Cori, B. Illingworth, and P. G. Keller in S. P. Colowick and N. O. Kaplan, *Methods in Enzymology* [New York: Academic Press, Inc., 1955], **1**, 200), the rest of the sample was used for the assay of incorporation.

Fractionation of the homogenate system was carried out by differential centrifugation. The results of such a fractionation are shown in Table 2. Since it had not been possible to substitute energy sources such as ATP, creatine P, or PEP for oxidative phosphorylation in the homogenate system, it was necessary, of course, to include mitochondria in the incorporating system. A surprising result, however, was the finding that the particle fraction precipitating between 20,000 × *g* and 100.000 × *g* seemed to be required for any sizable incorporation. An explanation of this effect appeared subsequently when, on determining the intracellular distribution of the UDPG glycogen synthesizing system, activity was found chiefly in the particle fraction (Table 4). Although indirect, this experiment provides preliminary evidence that incorporation in the homogenate system was probably by way of the UDPG system.

Synthesis of Glycogen from UDPG.—When C^{14}-labeled UDPG was tested with muscle homogenates, it was found to be incorporated much more rapidly than

TABLE 2*

DIFFERENTIAL CENTRIFUGATION OF PIGEON BREAST MUSCLE SYSTEM THAT INCORPORATES GLUCOSE INTO GLYCOGEN

Fraction	Glucose Incorporation (c.p.m.)
Mitochondria	< 100
Mitochondria + 20,000 × *g* supernatant	4,800
Mitochondria + 100,000 × *g* supernatant	600
Mitochondria + 100,000 × *g* supernatant + glycogen	500
20,000 × *g*. supernatant	< 100
100,000 × *g*. supernatant	< 100

* Pigeon breast muscle homogenate was prepared as described in the text but was centrifuged for 30 minutes at 20,000 × *g*. Part of the supernatant fraction was centrifuged for 60 minutes at 100,000 × *g* and the precipitate discarded. Mitochondria were isolated by differential centrifugation (J. W. Harman and U. H. Osborne, *J. Exptl. Med.*, **98**, 81, 1953) after resuspension and dilution of the 20,000 × *g* precipitate. Incubations were carried out with 0.1 ml. of mitochondrial suspension and 0.6 ml. of the appropriate supernatant fraction in a final volume of 1 ml., and 0.2 ml. of kochsaft was added to stimulate incorporation. Glycogen was added to a final concentration of 0.2 per cent where indicated. After 1 hour at 37° in O_2, glycogen was isolated and counted as described in the text.

glucose or glucose-6-phosphate. In fact, as shown in Table 3, when incubations were carried out under anerobic conditions with a small amount of dilute homogenate, incorporation was obtained only with UDPG. The maximum incorporation rate was found at pH 7.5, with 0.01 *M* Mg^{++}. Addition of primer glycogen caused an increase in the rate of incorporation. The reaction is approximately linear with enzyme concentration and with time. Paper electrophoresis confirms that the only product formed from UDPG is glycogen.

Homogenates or extracts prepared in the absence of EDTA had very little activity. The differential centrifugation of a rat muscle homogenate is shown in

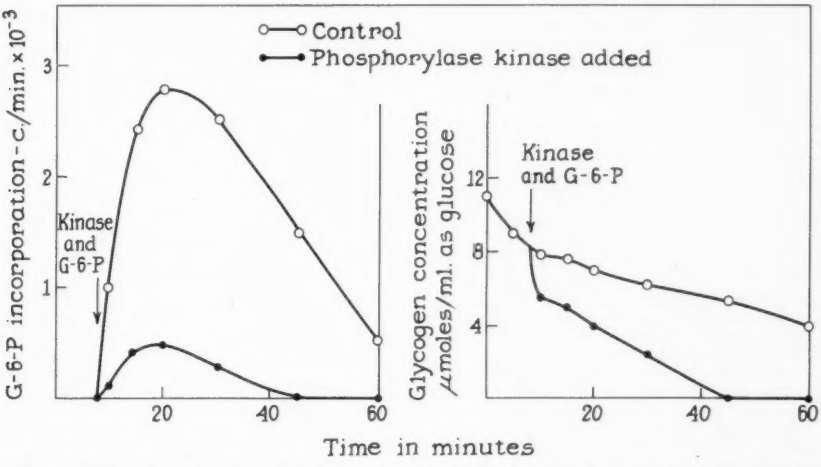

FIG. 2.—Phosphorylase kinase effect on glucose-6-phosphate incorporation into glycogen. The experiment is similar to that described in Fig. 1, except that two large-scale incubations were carried out in 125-ml. Erlenmeyer flasks. The glucose-6-phosphate concentration was 5×10^{-5} *M*, 100,000 cpm/ml. Samples were withdrawn at the times indicated for phosphorylase assays and incorporation determinations. For the chemical glycogen determinations, samples of the same isolated glycogen which was counted were hydrolyzed and assayed for reducing sugar, as described by Hassid and Abraham (in S. P. Colowick and N. O. Kaplan, *Methods in Enzymology* [New York: Academic Press, Inc., 1957], **3**, 34).

TABLE 3*

INCORPORATION OF GLUCOSE, GLUCOSE-6-PHOSPHATE, AND UDPG INTO GLYCOGEN BY RAT SKELETAL MUSCLE HOMOGENATE

Substrates	Incorporation (c.p.m.)
C^{14}-glucose, ATP, UTP	< 50
C^{14}-glucose-6-phosphate, UTP	< 50
C^{14}-UDPG	1,400

* The incubation medium contained 0.05 *M* glycylglycine, pH 7.5, 0.01 *M* $MgCl_2$, 1 per cent rabbit liver glycogen, and enzyme in a final volume of 0.5 ml. The reaction was started by adding the labeled substrate after a brief pre-incubation at 30°. Incubations were carried out for 15 minutes at 30°. The reaction was stopped by adding 0.5 ml. of 60 per cent KOH, and the tubes were mixed, covered, and heated in a boiling-water bath for 10 minutes. The rest of the procedure was carried out as described for the pigeon breast muscle assay system except that the volumes were half as large.

In this experiment, 0.02 ml. of a 1:20 homogenate of rat skeletal muscle in 0.25 *M* sucrose 10^{-2} *M* EDTA was used as the source of enzyme. The labeled compounds were all present at a concentration of 4×10^{-4} *M* and 10,000 c.p.m. Nucleotides were added to a final concentration of 10^{-3} *M*.

Table 4. All the activity remains in the supernatant fraction after 10 minutes at 20,000 × *g*, while the 100,000 × *g* supernatant fraction is inactive. The particles alone have some activity, but, as shown in Figure 3, the addition of the dialyzed supernatant fraction and inorganic phosphate causes substantial stimulation. Essentially, the same particle-supernatant system can be obtained from muscle minces by extracting at room temperature for 60 minutes with 0.01 *M* EDTA. Mince extracts made in the cold or in the absence of EDTA have little, if any, activity.

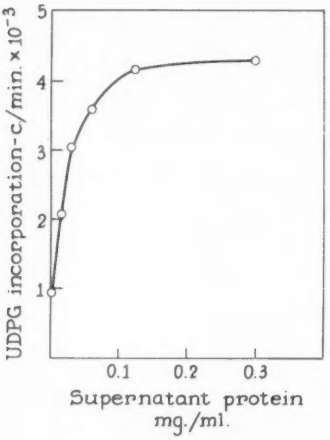

FIG. 3.—Stimulation of UDPG incorporation into glycogen by the dialyzed supernatant fraction from muscle. The particles used for this experiment were prepared from room-temperature extracts of rabbit muscle by acid precipitation. The Spinco supernatant fraction was dialyzed overnight against 0.01 *M* EDTA and gave no stimulation in the absence of added phosphate. It gave no incorporation by itself in the presence or absence of phosphate. Incorporation experiments were carried out as described in Table 3, except that 4×10^{-4} *M* K_2HPO_4 was added to all tubes. The dialyzed supernatant fraction was diluted with cold water before being used.

COMMENTS

Many of the chemical and biological aspects of glycogen synthesis from UDPG remain obscure at the present. It seems likely, however, that transfer of glucose to a glycogen primer leads to the formation of a 1,4-linked polymer which undergoes branching under the influence of amylo-(1,4 → 1,6)-transglucosidase.[15] Preliminary observations have shown that when UDPG is incubated with the particles alone, a polysaccharide is formed which gives a purple color with iodine, indicating that only partial branching has taken place with the particles and UDPG alone.

The reason that the particle reaction is stimulated by the addition of supernatant protein is not yet clear. The undialyzed supernatant fraction stimulates, but after

dialysis no stimulation was found unless inorganic phosphate was added. This observation suggested the need to check phosphorylase involvement in the reaction. It was found, however, that crystalline phosphorylase had little effect on the incorporation of UDPG by the particles. To exclude a nonspecific effect, serum albumin was added and found inactive. Experiments are being continued at present to define the activity of the supernatant fraction.

TABLE 4*

DIFFERENTIAL CENTRIFUGATION OF UDPG-INCORPORATING SYSTEM OF SKELETAL MUSCLE

Fraction	Incorporation (c.p.m.)
Whole homogenate	4,500
20,000 × *g* supernatant fluid	4,050
100,000 × *g* supernatant fluid	150
100,000 × *g* precipitate	1,200

* Rat skeletal muscle was homogenized in 10 volumes 0.25 *M* sucrose 10^{-2} *M* EDTA and was centrifuged successively for 10 minutes at 20,000 × *g* and 60 minutes at 100,000 × *g*. The 100,000 × *g* precipitate was resuspended in the original volume of sucrose EDTA. Incubations were carried out with 0.025 ml. of the appropriate fraction in a final volume of 0.5 ml. Assays were carried out as described in Table 3.

Preliminary experiments have shown that the activity in the 20,000–100,000 × *g* particle fraction is firmly bound. Since this fraction contains a high concentration of glycogen, the possibility that the enzyme is strongly adsorbed to glycogen must still be considered. The fact that almost half of the phosphorylase *b* of a muscle homogenate is also found in this fraction strengthens this possibility.

The thermodynamic advantages of glycogen synthesis by way of UDPG are obvious. Calculations based on data for the sucrose-synthesizing enzyme, sucrose phosphorylase, and phosphorylase[16, 17] show that the transfer of a glucose unit from UDPG to glycogen should have a standard free-energy change of about −3,000 cal. Conversion to glycogen, therefore, should be greater than 99 per cent. In confirmation, transfers of this order of magnitude have already been observed.

SUMMARY

In a pigeon breast muscle homogenate, the synthesis of glycogen was studied by following the incorporation of radioactive glucose and glucose-6-phosphate. Incorporation depended on active respiration, which was not replaced by ATP or energy-rich phosphate feeder systems. Maintenance of respiration in this prepation depended on the addition of muscle kochsaft, which could be almost completely replaced by the addition of CoA and aspartic acid.

Addition of phosphorylase kinase increased the concentration of phosphorylase *a* in the system and strongly inhibited the incorporation of glucose and glucose-6-phosphate. Insulin had no effect; adrenalin inhibited in high concentrations.

Glycogen synthesis from UDPG was found to be highly concentrated in particles which had microsomal sedimentation characteristics. These had to be supplemented for optimal activity by a nondialyzable supernatant factor, Mg^{++}, and phosphate.

The probable function of the UDPG enzyme as the cellular system for glycogen synthesis is discussed.

* This investigation was supported in part by a research grant from the National Cancer Institute, National Institutes of Health, U.S. Public Health Service, and the National Science Foundation.

The following abbreviations have been used: UDPG, uridine diphosphate glucose; ATP, adenosine triphosphate; DPN, diphosphopyridine nucleotide; TPN, triphosphopyridine nucleotide; CoA, coenzyme A; EDTA, ethylene diamine tetraacetic acid; UTP, uridine triphosphate; creatine P, creatine phosphate; PEP, phosphoenol pyruvate.

[1] D. M. Kipnis, and C. F. Cori, *J. Biol. Chem.*, **224,** 681, 1957.

[2] H. A. Krebs, and L. V. Eggleston, *Biochem. J.*, **32,** 513, 1938.

[3] F. J. Stadie, J. A. Zapp, Jr., and F. D. W. Lukens, *J. Biol. Chem.*, **132,** 453, 1940.

[4] E. W. Sutherland, Jr., in O. H. Gaebler (ed.), *Enzymes: Units of Biological Structure and Function* (New York: Academic Press, Inc., 1956), p. 536.

[5] T. W. Rall, E. W. Sutherland, and J. Berthet, *J. Biol. Chem.*, **224,** 463, 1957.

[6] T. W. Rall, and E. W. Sutherland, *J. Biol. Chem.*, **232,** 1065, 1958.

[7] G. T. Cori, and B. Illingworth, *Biochim. et Biophys. Acta*, **21,** 105, 1956.

[8] L. F. Leloir and C. E. Cardini, *J. Am. Chem. Soc.*, **79,** 6340, 1957.

[9] L. Glaser, and D. H. Brown, these PROCEEDINGS, **41,** 253, 1955.

[10] W. Z. Hassid, and S. Abraham, in S. P. Colowick, and N. O. Kaplan, *Methods in Enzymology* (New York: Academic Press, Inc., 1957), **3,** 34.

[11] O. Walaas, and E. Walaas, *J. Biol. Chem.*, **187,** 769, 1950.

[12] I. Lieberman, *J. Biol. Chem.*, **223,** 327, 1956.

[13] E. G. Krebs, and E. H. Fischer, *Biochim. et Biophys. Acta*, **20,** 150, 1956.

[14] E. H. Fischer, and E. G. Krebs, *J. Biol. Chem.*, **231,** 65, 1958.

[15] J. Larner, *J. Biol. Chem.*, **202,** 491, 1953.

[16] C. E. Cardini, L. F. Leloir, and J. Chiriboga, *J. Biol. Chem.*, **214,** 149, 1955.

[17] C. F. Cori, G. T. Cori, and A. A. Green, *J. Biol. Chem.*, **151,** 39, 1943.

*DISSOCIATION OF HUMAN CARBONMONOXYHEMOGLOBIN AT HIGH pH**

BY ULRICH HASSERODT AND JEROME VINOGRAD

GATES AND CRELLIN LABORATORIES OF CHEMISTRY,† CALIFORNIA INSTITUTE OF TECHNOLOGY, PASADENA, CALIFORNIA

Communicated by Linus Pauling, February 20, 1958

This communication presents a study of some of the hydrodynamic properties of the half-molecules of human carbonmonoxyhemoglobin at pH 11.0 and ionic strength 0.25, together with a demonstration that, at this pH, human carbonmonoxyhemoglobin is completely dissociated and the molecular weight of the product is one-half that of the whole molecule. The half-molecules appear to be stable when stored at pH 11 and 3° C., in contrast to their behavior in an acid environment, pH 4.5–6, in which denaturation occurs. The acid denaturation is so rapid at pH values low enough to induce complete dissociation that it has not been possible until now to study the properties of the half-molecules uncomplicated by the presence of the whole molecules. Results of a comparison of some properties of untreated hemoglobin and hemoglobin reconstituted from half-molecules by successive dialyses at pH 11 and pH 7 are also presented.

Experimental.—Crystallized carbonmonoxyhemoglobin was prepared according to the method of Drabkin.[1] This material was used in most of the sedimentation

studies. Hemoglobin obtained prior to crystallization was also employed, and no significant difference was found in the sedimentation behavior at either pH 7.0 or pH 11.0.

The following buffer solutions were used: Tris-HCl-NaCl, pH 8–9; glycine NaOH-NaCl, pH 9.6–11.6; and phosphate-NaCl, pH 11.0–11.6. The ionic strength was 0.25 in all cases. All hemoglobin solutions were dialyzed at 3° C. for 12–15 hours before the sedimentation runs.

The sedimentation and diffusion measurements were carried out in the Model E Spinco analytical ultracentrifuge, using a phase-plate schlieren diaphragm. A plastic commercial interference synthetic boundary cell[2] was used for the diffusion measurements, which were performed with an essentially static boundary at 8,000 rpm. Sedimentation velocity measurements in the pH range 10.0–11.6 were carried out in a double-cell rotor, using a 1° positive wedge. The second cell contained

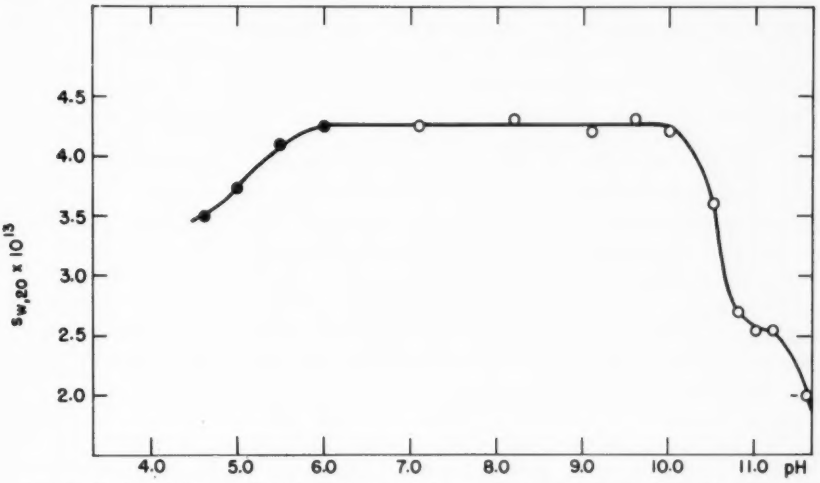

FIG. 1.—Sedimentation coefficients of human carbonmonoxyhemoglobin in the pH range 4.5–11.6. ● from Field and O'Brien (*Biochem. J.*, **60**, 656, 1955): Hb concentration: 0.71 per cent. Ionic strength of the solution: 0.1. ○ this study.

carbonmonoxyhemoglobin at pH 11.0 as a reference. Thus differences of 0.1 sedimentation units could be detected with considerable certainty.

The sedimentation and diffusion coefficients are expressed in the text in c.g.s. units $\times 10^{13}$ and 10^{7}, respectively.

Results and Discussion.—Table 1 shows the average sedimentation coefficients obtained in basic solutions. The trend of the values indicates that the dissociation of the hemoglobin molecule in a 1 per cent solution begins just beyond pH 10. At pH 11.0 the value of $s_{20,w}$ is 2.55. This is the value reported for horse globin[3] and also for human globin[4] at pH 7.0 and at concentrations of 1 per cent. Both these materials have half the molecular weight of the corresponding hemoglobins.[3,4] Figure 1 shows a small plateau in the s versus pH curve around pH 11.0, indicating that in this region hemoglobin is completely dissociated into two subunits. At this pH the dependence on concentration is normal, as would be expected for a fully

dissociated hemoglobin, while at pH 10.5 the results suggest the presence of incomplete dissociation (Table 1).

Molecular weight determinations at pH 11.0, using the Archibald method of approached sedimentation equilibrium,[5] gave an average value of 36,000 ± 2,000.[6] Involved in this result is the assumption that the partial specific volume does not change significantly from 0.749 ml. g.$^{-1}$.

The spreading of the boundaries at pH 10.5 and 11.0 is in accord with the above conclusions. The sedimenting boundaries after 1 hour or longer at 56,100 rpm were skewed at pH 10.5, and the spreading was greater than could be accounted for by diffusion alone. At pH 11.0 the boundaries had an approximately Gaussian shape, with less spreading than at pH 10.5 or 11.6. Thus, in plots of the standard deviation of the schlieren patterns σ versus $\sqrt{t}$, curves with an upward trend were obtained for Hb at pH 10.5 and 11.6, and a straight line at pH 11.0.

A more extended study at pH 11.0 gave the following results: the average diffusion coefficient (static) in phosphate as well as glycine-NaOH buffers is 7.6 ± 0.4; the concentration dependence lies within the limits of error; $s^\circ_{20,w}$ has the value 2.7, and $D^\circ_{20,w}$, calculated from s and M, has the value 7.4.

Moving-boundary electrophoresis at pH 11.0, $\mu = 0.1$, phosphate buffer, showed a single symmetrical boundary in the ascending and descending limb. A solution first dialyzed to pH 11 and then to pH 6.75 gave electrophoretic patterns indistinguishable from an equivalent sample dialyzed to pH 6.75 directly. A mixture of equal quantities of the foregoing materials again gave the same pattern.

The frictional ratio calculated for the subunit at pH 11.0 is 1.27, whereas this value for hemoglobin at pH 7.1 is 1.21, if $s^\circ_{20,w}$ is taken to be 4.45 and the molecular weight 66,500.[7] This slight increase in the molecular asymmetry, which corresponds, for example, to a change in the axial ratio from 3:1 to 4:1 in an ellipsoid, assuming 20 per cent hydration,[8] suggests that cleavage is along the major axis of the molecule.

The dissociation into subunits could be reversed upon neutralization. Samples of hemoglobin which had been at pH 11.0 for 12 and 36 hours, respectively, were dialyzed to pH 8.2, and the sedimentation coefficient was measured along with reference samples at the same pH. The result was 4.25.

Neither precipitation nor change in the spectrum (350–650 mμ) was observed on raising the pH to 11.0 and maintaining this pH for 36 hours. The region of the hemoglobin spectrum most sensitive to denaturation and shifting of the hemes is to be found at the maxima at 540 and 568 mμ. The ratios of the minimum at 376 mμ to the two maxima were 1.26 and 1.29 at pH 7.1 and 1.25 and 1.28 for Hb which had been at pH 11.0 for 36 hours and was redialyzed to pH 7.1.

On the other hand, the changes in the hemoglobin molecule at pH 11.6 are not reversible. The color of the solution changes, and the maxima at 540 and 568 mμ disappear slowly. Precipitation occurs on redialysis to pH 7.1, and the amount of this precipitate increases with the time of exposure to the alkali.

From the data in Table 1 and the results of Field and O'Brien[9] a pH stability diagram may be formulated (Fig. 1). This curve is not symmetrical about the isoelectric point, indicating that dissociation is not entirely the result of repulsions generated by the net negative or positive charge on the protein. Neglecting ion binding, it may be estimated[10] that at pH 6, where acid dissociation begins, the

TABLE 1

SEDIMENTATION AND DIFFUSION COEFFICIENTS OF HUMAN CARBONMONOXYHEMOGLOBIN*

pH	1 Per Cent $s_{20,w}$	0.3 Per Cent $s_{20,w}$	$D_{20,w}$
8.2	4.3	...	...
9.1	4.2	...	...
9.6	4.3		
10.0	4.2†	4.35	...
10.5	3.6†	3.4	...
10.8	2.70†	...	...
11.0	2.55	2.65	7.6 ± 0.4
11.2	2.55†	...	...
11.6	2.0†	...	...

* All measurements at $\mu = 0.25$ in buffered solutions described in text.
† Runs in double-cell rotor with a solution at pH 11 present as a reference.

molecule has a net charge of +5, whereas at pH 10 the charge is about −30. These considerations neglect binding of negative ions, which may be expected to make the situation even more asymmetric. Thus it is not likely that general net charge repulsions are solely responsible for the dissociation reaction at low pH.

An examination of the stability curve suggests that the amino acid residues critically involved in stabilizing the whole molecules are histidines, which are protonated during acid dissociation, and lysines and tyrosines, which are neutralized or ionized upon alkaline dissociation.

The temperature coefficients of the dissociation into subunits have been estimated from the temperature coefficients of the observed sedimentation coefficients at pH 5 and 10.5. The results corresponded to 0.0 ± 0.5 kcal./mole at pH 5 and -4.0 ± 0.5 kcal./mole at 10.5. These enthalpy changes are interpreted as the sums of the enthalpies of splitting the molecules plus the enthalpy changes involved in charging or discharging the critical groups. Since the number of the groups is not known, it is not possible at the present time to apportion these enthalpy changes.

* This work was supported in part by a grant No. H-8394 from the United States Public Health Service.

† Contribution No. 2289.

[1] D. L. Drabkin, *Arch. Biochem. and Biophys.*, **21**, 224, 1949.

[2] Spinco Division, Beckman Instruments Co.

[3] N. Gralen, *Biochem. J.*, **33**, 1907, 1939.

[4] U. Hasserodt, unpublished.

[5] S. M. Klainer and G. Kegeles, *J. Phys. Chem.*, **59**, 952, 1955.

[6] This value is not corrected for the presence of a non-heme protein, which is present to the extent of 1–2 per cent even in crystallized samples of human hemoglobin (D. W. Allen, W. A. Schroeder, and J. Balog, *J. Am. Chem. Soc.*, **80**, 1628, 1957). This protein, which has a different amino acid composition from globin, has a sedimentation coefficient of 8.6 and an estimated molecular weight of 180,000. The correction for its contribution to the weight average molecular weight leads to a value of 34,000 for human hemoglobin at pH 11.0.

[7] U. Hasserodt, J. Vinograd, and R. Srinivasan, *J. Am. Chem. Soc.* (in press).

[8] J. L. Oncley, *Ann. N.Y. Acad. Sci.*, **41**, 121, 1941.

[9] E. O. Field and J. R. P. O'Brien, *Biochem. J.*, **60**, 656, 1955.

[10] E. J. Cohn, A. A. Green, and M. H. Blanchard, *J. Am. Chem. Soc.*, **59**, 509, 1937.

RADIATION INDUCED REVERSE MUTATIONS IN DROSOPHILA MELANOGASTER

By M. M. Green

DEPARTMENT OF GENETICS, UNIVERSITY OF CALIFORNIA, DAVIS

Communicated by G. Ledyard Stebbins, November 7, 1958

There can be little doubt that within the limits of current operational criteria spontaneous reverse (back) mutations have been recorded in *Drosophila* and other organisms.[1] Doubt has been raised whether the back-mutation frequency can be significantly increased following X-ray irradiation, despite early reports that it can.[2]

The relatively high spontaneous back-mutation frequencies of two pseudoallelic, recessive, sex-linked forked bristle (*f*) mutants in *Drosophila melanogaster* suggested that they might provide critical information bearing on the question of induced reverse mutation following X-ray or gamma treatments. These mutants, f^{3n} and f^{1}, were found to back-mutate spontaneously in females at rates of $46.7/10^6$ (14/298,698) and $15.3/10^6$ (7/457,316) chromosomes, respectively.[1]

By design, attached-X ♀ ♀ of genotypes *v* f^{3n}*car* and *v* f^{1}*B* were treated with 4,000 r X-rays or gamma, and reversals were sought among their female progeny. The use of attached-X ♀ ♀ permits the scoring of two treated chromosomes for each ♀ zygote.

Adult females 0–48 hours old were irradiated with either 4000 r X-rays delivered from a General Electric "Maxitron" machine operated at 30 m.a. and 100 KVP (rate ca. 1,000 r/min.) or 4,000 r gamma delivered from a Co^{60} source (rate 265,000 r/hour).[3] Treated ♀ ♀ following mating to f^{1} ♂♂ were allowed to oviposit for 7 days and transferred to fresh media for a second 7-day interval. Partial and complete *f* reversals were sought among F_1 ♀ ♀. All ♀ ♀ showing phenotypic reversals were progeny-tested by crossing to ♂ sibs. Progeny tests were obtained on all ♀ ♀ discussed herein.

The results of these experiments, listed in Table 1, represent summations of three separate X-ray treatments of *v* f^{3n}*car* ♀ ♀ and two of *v* f^{1}*B* ♀ ♀, together with two separate gamma treatments of *v* f^{3n}*car* ♀ ♀ and one *v* f^{1}*B* ♀ ♀. The reversals recorded were all complete and phenotypically inseparable from wild-type. All bred true on progeny-testing. Although a number of partial reversals were recorded, none bred true.

TABLE 1

COMPLETE REVERSALS OF f^{3n} AND f^{1} IN ATTACHED-X ♀ ♀ FOLLOWING X-RAY OR GAMMA IRRADIATION

Radiation (4,000 r)	Genotype Treated ♀ ♀	Total *f*⁺ ♀ ♀	Total Chromosomes	Reversals/10^6 Chromosomes
X-rays	*v* f^{3n} *car*	38	115,774	328.2
	v f^{1}*B*	1	68,268	14.6
Gamma	*v* f^{3n}*car*	10	65,206	153.3
	v f^{1}*B*	0	21,550	0

A comparison of the induced and spontaneous rates of reversals shows that 4,000 r X-rays increased f^{3n} reversals seven fold, whereas equivalent gamma irradiation somewhat more than tripled the reversal frequency. There is little doubt that radiation at the dose given significantly increased the reversal frequency.

The experiments with f^1 show that no comparable significant increase in reversal frequency followed irradiation. Admittedly, the less extensive nature of these experiments permit only a tentative conclusion.

Incidental to scoring *f* reversions, flies were also scored for possible reversions of *v*, *car*, and *B*. No reversals of *v* or *car* were found. At least one partial reversal of *B* was found, and this is not unexpected, since induced partial and full reversions of *B* have been recorded.

By progeny-testing reversal ♀ ♀, two relevant questions can be answered: (1) whether one or both treated *f* mutants reverted following irradiation and (2) whether gross chromosomal rearrangements are causally associated with reversion. The occurrence of crossing over in attached-X ♀ ♀ provides answers to both questions. If but one *f* mutant reverted following irradiation, then the unaltered *f* should be recovered in homozygous condition as a consequence of nonreciprocal crossing over between the *f* locus and the centromere. If no gross chromosomal rearrangement occurred, the frequency of homozygosis of the unaltered *f* mutant should equal one-fourth the standard frequency of crossing over between *f* and the centromere.

Without presenting the extensive data, the following conclusions can be made. First, homozygous f^{3n} or f^1 ♀ ♀ were recovered among the progeny of all reversals, proving that only one of the two treated *f* alleles reverted. Second, in each case the frequency of homozygosis was consistent with the interpretation that crossing over between the *f* locus and the centromere was unencumbered by any gross chromosomal rearrangement.

Ample evidence has accumulated to warrant the conclusion that *Drosophila* cells in different stages of spermatogenesis differ significantly in their sensitivity to X-irradiation.[4] A similar situation during oögenesis is indicated by studies of induced dominant and recessive lethals.[5] Following X-irradiation, the frequency of dominant and recessive lethal drops sharply after the seventh day of oviposition, presumably because the supply of mature oöcytes exposed at the time of treatment have been replaced by cells that were younger (oögonia) at that time. Since dominant lethals are, in all probability, associated with chromosomal rearrangements, it follows that the maturer cells are more prone to induction of rearrangements than are younger cells. Hence a comparison between reversion rates during the two oviposition periods could demonstrate whether or not reversions are associated with chromosomal rearrangements.

A consideration of the f^{3n} reversions induced by X-rays shows that the rates for the first and second weekly egg-laying intervals are essentially identical. For the first week, 8 reversions/29,452 chromosomes were recorded; for the second week, 30/86,322. The rates computed per 10^6 chromosomes are 271.6 versus 347.5 and appear not to be significantly different. This comparison supports the conclusion stated earlier that the reversions appear to be independent of chromosome rearrangement. If such an association exists, a preponderance of reversions would be expected among the progeny of the first 7-day egg-laying interval. The data suggest that perhaps most, if not all, of the reversions are induced in young oöcytes or oögonia. Among *D. melanogaster* females 0–48 hours old, the majority of ovarian cells at the time of treatment will be at the oögonial or early oöcyte stage of development. Exposure to 4,000 r X-rays or gamma should induce dominant lethals in most mature oöcytes, thereby leaving only the viable oögonia and young oöcytes

eligible for reversions. If the latter cells are the primary source of reversions then, as observed, no difference is expected in rate between the consecutive egg-laying intervals.

The question of whether the phenotypic reversion of recessive mutants to wild type or near wild type is in fact back mutation has been discussed *in extenso* elsewhere in connection with the problem of spontaneous back mutation.[1] The arguments militating for spontaneous back mutation as a specific change or alteration of a particular mutant gene apply equally well for the radiation-induced reversions of f^{3n} reported here and need no further comment. In short, these results amplify results based primarily on partial reversions in *Drosophila*[6] and of reversions in *Neurospora*[7] that X-rays and gamma rays can induce back mutations which, insofar as current criteria permit, do not differ from spontaneous back mutations.

Summary.—*Drosophila melanogaster* ♀ ♀ homozygous for one or the other forked pseudoalleles, f^{3n} and f^{1}, both known to back-mutate spontaneously, were irradiated with 4,000 r X-rays or gamma rays.

Back mutations of f^{3n} were significantly increased; those of f^{1} apparently not.

The back mutations are not associated with detectable chromosome alterations.

[1] M. M. Green, *Genetica*, **29**, 1–38, 1957.

[2] G. Lefevre, *Am. Naturalist*, **84**, 341–365, 1950.

[3] Gamma irradiation experiments were carried out at the Biology Department, Brookhaven National Laboratory. Thanks are due Drs. H. J. Curtis and R. C. King for making facilities available.

[4] C. Auerbach, *Z. f. Vererbungs.*, **86**, 113–125, 1954; A. F. E. Khishin, *Z. f. Vererbungs.*, **87**, 97–112, 1955; K. G. Lüning, *Hereditas*, **38**, 91–107, 1952; I. I. Oster, *Adv. Radiobiology*, pp. 475–480, 1957.

[5] R. C. King, J. B. Darrow, and N. W. Kaye, *Genetics*, **41**, 890–900, 1956.

[6] H. J. Muller, and I. I. Oster, *Adv. Radiobiology*, pp. 407–413, 1957.

[7] N. H. Giles, F. J. deSerres, and C. W. H. Partridge, *Ann. N.Y. Acad. Sci.*, **59**, 536–552, 1955; F. J. deSerres, *Genetics*, **43**, 187–206, 1958.

DEPENDENCE OF MUTATION FREQUENCY ON RADIATION DOSE RATE IN FEMALE MICE

BY W. L. RUSSELL, LIANE BRAUCH RUSSELL, AND MARY B. CUPP

BIOLOGY DIVISION, OAK RIDGE NATIONAL LABORATORY,* OAK RIDGE, TENNESSEE

Communicated by Alexander Hollaender, November 17, 1958

In studies of specific locus mutation rates in mouse spermatogonia, a much lower mutation rate was obtained from chronic gamma than from acute X-irradiation.[1–4] Differences in both intensity (dose rate) and quality of radiation were involved in the comparison, but it was pointed out[2] that, although gamma rays frequently show a biological effectiveness somewhat lower than that of X-rays, the observed difference between the mutation rates appeared to be greater than could be attributed to this factor. This point of view has since been confirmed by three separate experiments conducted to test it.[4] In these experiments, no appreciable differences were found in the effectiveness of *acute* gamma rays (from Co^{60}), on the one hand, and acute X-rays, on the other, in inducing dominant lethals in spermatozoa, specific

locus mutations in spermatozoa and other postspermatogonial stages, or specific locus mutations in spermatogonia. It can be concluded that most of the difference between the *chronic* gamma and acute X-ray mutation rate results must be due to dose rate.

However, the finding of a dose-rate effect in spermatogonia did not necessarily prove that this effect was on the mutation process itself. It might, instead, have been due to secondary causes, such as cell selection resulting from spermatogonial killing or from other interference with the dynamics of the cycle of the seminiferous epithelium.[1, 2] To determine whether the mutation process itself could be affected by dose rate, it became important to carry out a dose-rate comparison in females where there seems little possibility that mutation-rate measurements can be interfered with by such secondary processes as elimination of sensitive cells. The adult ovary contains no oögonia, and the population of germ cells consists solely of oöcytes, which are non-mitotic and which appear to be remarkably uniform.[5] Furthermore, the continued fertility of females in chronic radiation experiments[6, 7] provides no evidence of extensive killing, selective or otherwise, of the oöcytes.

The first part of the dose-rate comparison for mutation rates induced in oöcytes has already been reported. The specific locus mutation rate in adult females exposed to continuous chronic Cs^{137} gamma irradiation at a dose rate of 86 r per week proved to be considerably lower than that produced by acute X-rays in spermatogonia.[6] Carter[8] also observed a low mutation rate in females exposed to interrupted chronic Co^{60} gamma irradiation of somewhat higher intensity than that used by us. The mutation rates from chronic irradiation of females, as far as they go, are not significantly different from the rates obtained from chronic gamma irradiation of spermatogonia.[4]

The completion of the dose-rate comparison for mutation rates induced in oöcytes, namely, the determination of mutation rates in females exposed to acute irradiation, promised a clear-cut answer of broad significance. For, if this rate proved to be no higher than that from chronic irradiation, then it would show that oöcytes are much less sensitive than spermatogonia. On the other hand, if the rate from acute irradiation of females turned out to be similar to that from acute irradiation of spermatogonia, namely, much higher than from chronic irradiation of both sexes, then it would indicate that there was a dose-rate effect directly on the mutation process.

It was, therefore, most important to obtain a reliable mutation rate from acutely irradiated females, even though this task is beset by the practical difficulty that the number of offspring that can be obtained before sterility ensues is very small. The experiment reported here was accordingly started to extend the limited data available from an earlier pilot experiment.[6] A brief statement of the new findings has already been presented.[4] In the present account, the results to date are described in more detail. Although the collection of data is still proceeding, the results are already sufficient to give a clear answer to the questions posed.

Materials and Method.—Young adult F_1 hybrid females from a cross of 101 × C3H strains were exposed to 400 r of whole-body X-radiation (250 Kvp; 15 ma.; inherent filtration 3 mm. Al; H.V.L. 0.4 mm. Cu; dose rate 92 r/min; animals in lucite containers on turntable rotating above masonite scatter block). The females were mated for 2 weeks, beginning on the day following irradiation, to

males of our T stock, which is homozygous for seven recessive marker genes. Young subsequently born were examined for mutations at the seven loci. Details of the experimental procedure have been described earlier.[9]

The data obtained in the present, 1958, experiment are additive with those of the 1950–1951 series,[6] since the stocks used were the same and the methods were identical except for one detail: in 1950–1951, when fertility effects of radiation were still being explored, females were kept with their mates indefinitely instead of being separated after 2 weeks, as in the 1958 series. However, the bulk of the young (97 per cent) were conceived within 2 weeks after irradiation, since most females were sterile by the time they could have had a second mating, and all were sterile 5 weeks after irradiation.

In 1950–1951, 379 females were irradiated, and 319 of these gave 1,729 young surviving to the age at which mutations are scored (i.e. approximately 3 weeks). In 1958, 6,130 young have, so far, been scored for mutations. These were offspring of 1,088 irradiated females. The results of the two acute X-ray experiments are presented in Table 1, together with data from chronic irradiation of females and control data. The data from the chronic gamma experiment have been brought up-to-date since the last report.[6]

TABLE 1

MUTATIONS AT SPECIFIC LOCI IN MOUSE OÖCYTES

RADIATION	DOSE (r)	INTENSITY (r/min)	NO. OF OFFSPRING	NO. OF MUTATIONS AT LOCUS							
				a	*b*	*c*	*d*	*p*	*s*	*se*	*d se*
—*	0	—	5,845	—	—	—	—	—	—	—	—
X*	400	96	1,729	—	—	—	—	—	1	—	—
X	400	92	6,130	—	2	—	—	1	2	—	1
—	0	—	40,918	—	—	—	—	—	—	—	—
Gamma (Cs^{137})	258	0.0085	26,468	—	1	—	—	—	—	—	1

* 1950–1951 experiment (Russell, Russell, Gower, and Maddus, these PROCEEDINGS, **44**, 901–905, 1958).

Conclusions.—On combining the 1950–1951 and the 1958 series on acute irradiation of oöcytes, the mutation rate is 31.8×10^{-8}/r/locus (95 per cent confidence interval: $12.8–65.5 \times 10^{-8}$). The mutation rate calculated from the up-to-date tabulation of the chronic gamma data for females is 4.2×10^{-8}/r/locus (95 per cent confidence interval: $0.51–15.1 \times 10^{-8}$). A test of the significance of the difference between the mutation rates per roentgen in the two experiments gave $P = 0.006$ for a one-tailed test. Thus a significantly greater frequency of specific locus mutations is induced in oöcytes by acute than by chronic irradiation.

It might be argued that the oöcyte stages irradiated were not strictly comparable in the chronic and acute radiation experiments. Thus, in the former, offspring were obtained for almost a year, representing, therefore, oöcytes irradiated in follicles of a variety of stages of development, from youngest to most mature. In the acute irradiation experiment, on the other hand, all offspring resulted from oöcytes irradiated in relatively mature follicles. However, as already noted, the state of the oöcyte nucleus in the adult ovary appears quite uniform (dictyate) throughout the development of the follicle, except for the last few hours before ovulation (a stage avoided in the present experiment by not mating females until 4 day after irradiation). Development of the follicle seems to involve only growth in oöcyte size and multiplication of follicle cells. Moreover, a difference between

chronic and acute irradiation can be demonstrated even if one uses only that part of the chronic radiation data which pertains to oöcyte stages more comparable to those involved in the acute radiation experiment, i.e., only litters conceived within 2 weeks following accumulation of the dose. In Carter's experiment,[8] in which all litters were of this classification (because of subsequent sterility from the dose distribution used), one mutation occurred among 10,117 offspring. Of the 26,468 offspring from chronically irradiated females in our experiments, 4,303 were conceived in the first 2 weeks, and one of the two mutations was among these. These two sets of results from chronically irradiated later oöcyte stages, taken together, give a significantly lower rate than that now found to result from acutely irradiated females ($P = 0.005$, for one-tailed test), indicating that, as suspected, the difference between the acute and the total chronic data is not due to oöcyte stage.

The finding of an effect of dose rate on the frequency of mutations recovered from irradiated females—where, as explained, secondary factors, such as cell selection, cannot be easily invoked—argues strongly that the mutation process itself must be affected by dose rate. Furthermore, although secondary mechanisms cannot yet be definitely ruled out as the cause of the intensity effect demonstrated earlier for spermatogonia, it is simpler to assume that the explanation for the results in oöcytes also applies to spermatogonia.

That the frequency of radiation-induced mutations is independent of dose rate has been one of the basic tenets of radiation genetics. The finding that dose rate affects the frequency of specific locus mutations observed following irradiation of mouse spermatogonia[1-3] did not by itself violate this tenet, since the result might have been due to secondary factors. However, the present demonstration of a dose-rate effect on mutation rate in oöcytes shows that the tenet does not have universal validity. The classic findings of intensity independence were made in *Drosophila* spermatozoa.[10] Recent experiments on mouse spermatozoa[4] are in complete agreement with the *Drosophila* results. Therefore, the explanation for the new phenomenon of intensity dependence resides in gametogenic stage. Hypotheses concerning the mechanism for this effect have been discussed elsewhere,[4] and it was shown that a threshold response for all mutations in spermatogonia and oöcytes is not a necessary consequence of the findings.

The specific locus mutation rates in oöcytes, 4.2×10^{-8}/r/locus and 31.8×10^{-8}/r/locus for chronic and acute irradiation, respectively, may be compared with the rates in spermatogonia, namely (as calculated from straight lines fitted in the range of 0–861 r[4]), 5.2×10^{-8}/r/locus and 21.3×10^{-8}/r/locus for chronic and acute irradiation, respectively. It will be noted that, at each dose rate tested, there is, at present, no evidence of marked difference between oöcytes and spermatogonia. Before the mutation rates from acute irradiation of females were known, it was argued,[6] on the basis of the results then available, that the low mutation rate obtained with chronic irradiation was due either (*a*) to the same factor in males and females, namely, the low dose rate, or (*b*) to different factors in the two sexes, namely, to secondary causes in males and to a low sensitivity of the sex or of the oöcyte stage in females. The new results, showing a high mutation rate from acute irradiation of females, obviously disprove the second hypothesis. The conclusion,[6] that "the mutation rate in women exposed to chronic gamma radiation may be less than that estimated from our earlier yardstick, namely, the mutation

rate obtained from acute X-irradiation of male mice," can now be extended to chronically exposed *men* also, since the new results have made secondary factors unlikely. Carter[8] thought it most likely that the low mutation rate found following chronic irradiation of oöcytes was attributable to sex. As has just been pointed out, this interpretation is not upheld by the results reported in this paper. Carter's emphasis on the consequence of his interpretation, namely, that only a small part of the genetic hazard from medical irradiation would come from exposure of females, now appears to have been misleading.

Of course, more information is needed on the relative mutational sensitivities of males and females under various conditions of irradiation—intensity, fractionation of dose, etc.—but there is no justification at present for assuming less genetic hazard from irradiation of females. In fact, the present point estimate from the mutation rate induced by acute radiation in oöcytes is actually slightly higher than the spermatogonia rate. Although the difference is not significant in the present data, the possibility that the mutation rate may be somewhat greater in oöcytes than in spermatogonia is not out of line with the higher mutation rates found from irradiation of postspermatogonial stages,[11] among which are spermatocytes. In *Drosophila*, specific locus mutation rates are about half to three-fourths as high in oöcytes as in spermatozoa,[12] but the spermatozoa rate is about four times the spermatogonia rate.[13] Thus the relative mutation rates of oöcytes and spermatogonia may be similar in the two species.

It may be noted from Table 1 that no mutations have, to date, occurred among 46,763 offspring of our control females. Control males have, so far, yielded 17 mutants among 288,616 young[4] at our laboratory and 4 mutants among 41,262 young at Harwell.[8] Thus there is some indication that the spontaneous mutation rate in females may be lower than in males in the mouse. A similar situation exists in *Drosophila*.[14]

Summary.—The specific locus mutation rate in females is significantly higher following acute than following chronic irradiation. Since conditions in the adult mouse ovary are such that complications from secondary factors are unlikely, it can be concluded that in oöcytes there is a dose-rate effect on the mutation process itself. The influence of dose rate on the frequency of mutations following irradiation of spermatogonia, which was demonstrated earlier and which could have been due to secondary factors, is now, in the light of the oöcyte results, also considered to be due to a dose-rate effect on the mutation process.

The most significant features of the results, from the point of view of human hazards, are, first, that there is, at present, no evidence that females are less sensitive than males and, second, that the finding of a dependence of mutation frequency on radiation dose rate, now extended from males to females, applies to those germ cell stages that are important in human hazards.

The authors are grateful to Miss E. M. Kelly and Mrs. K. F. Stelzner for their extensive assistance in this experiment.

* Operated by Union Carbide Nuclear Company for the United States Atomic Energy Commission.

[1] W. L. Russell and E. M. Kelly, *Science*, **127**, 1062, 1958.

[2] W. L. Russell and L. B. Russell, *Proc. Second Intern. Conf. on Peaceful Uses of Atomic Energy* (Geneva, 1958) (in press).

[3] W. L. Russell, L. B. Russell, and E. F. Oakberg, in *Radiation Biology and Medicine*, ed. W. D. Claus (Reading, Mass.: Addison-Wesley Publishing Co., Inc., 1958), chap. viii.

[4] W. L. Russell, L. B. Russell, and E. M. Kelly, *Science* (in press).

[5] E. F. Oakberg, *Proc. Tenth Intern. Congr. Genetics*, **2**, 207, 1958.

[6] W. L. Russell, L. B. Russell, J. S. Gower, and S. C. Maddux, these PROCEEDINGS, **44**, 901–905, 1958.

[7] L. B. Russell and M. K. Freeman, *Radiation Research*, **9**, 174–175, 1958.

[8] T. C. Carter, *Brit. J. Radiol.*, **31**, 407–411, 1958.

[9] W. L. Russell, *Cold Spring Harbor Symposia Quant. Biol.*, **16**, 327–336, 1951.

[10] H. J. Muller, in *Radiation Biology*, Vol. **1**, ed. A. Hollaender (New York, N.Y.: McGraw-Hill Book Co., Inc., 1954), chap. viii.

[11] W. L. Russell, J. W. Bangham, and J. S. Gower, *Proc. Tenth Intern. Congr. Genetics*, **2**, 245–246, 1958.

[12] B. Glass, *Brookhaven Symp. Biol.*, **8**, 148–167, 1955.

[13] M. L. Alexander, *Genetics*, **39**, 409–428, 1954.

[14] B. Glass and R. K. Ritterhoff, *Science*, **124**, 314–315, 1956.

---

*MODIFICATION OF IRRADIATION EFFECTS IN THE PIGEON, COLUMBA LIVIA. I. EFFECTS OF HOMOLOGOUS AND HETEROLOGOUS BONE MARROW TRANSPLANTS**

BY DONALD H. SHAW AND HALVOR VERMUND

DEPARTMENT OF GENETICS AND DEPARTMENT OF RADIOLOGY, UNIVERSITY OF WISCONSIN

Communicated by M. R. Irwin, November 24, 1958

INTRODUCTION

Previous investigations have demonstrated that the lethal effects of total-body roentgen irradiation on mammals (mice, rats, rabbits, and guinea pigs) can be modified by isologous, homologous, or heterologous bone marrow transplantation.[1,10,11] These modifications appear to depend upon (1) the establishment and proliferation of the hematopoietic elements of the transplant and (2) either partial or complete repopulation of the host's depleted hematopoietic system by the cells proliferating from the graft.[3,4,7,8]

The present report includes preliminary findings of experiments in which homologous and heterologous bone marrow injections were administered to pigeons previously subjected to 2,500 rads (approximately the $LD_{100/30}$) total-body X-irradiation. The evidence indicates strongly that partial and eventually total repopulation of the erythrocytes results from the homoglous and heterologous transplants. Also, host-species-specific hemagglutinins (hemagglutinins specific for the erythrocytes of host species) appeared in the circulation of irradiated pigeons which received heterologous bone marrow.

MATERIALS AND METHODS

Dosage response studies have shown that the $LD_{50/30}$ for the pigeon is approximately 2,050 rads and that the effective $LD_{100/30}$ was 2,500 rads. These studies also

indicated that the death pattern appears to be comparable to that observed in mice, rats, etc. This information will appear in detail elsewhere.

Non-inbred sexually mature pigeons, 6–10 months of age, were individually exposed to 2,500 rads total-body roentgen irradiation. A single dose was administered by a General Electric Maxitron resonant transformer X-ray machine operated at 1,000-KV peak. The target served as a filter, and no additional filtration was provided. The half-value layer was 3.4 mm. of lead. The birds were treated individually and rotated at 5 revolutions per minute in a lucite cylinder measuring 7.5 by 20 cm., with the longitudinal axis of the animal as the axis of rotation. The radiation field covered the entire cylinder with the bird in the middle. The distance between the target of the X-ray tube and the axis of rotation was 70 cm. The dose rate was 92 r/min, as determined with a Victoreen roentgen meter using an ionization chamber placed within the cylinder with tissue-equivalent scattering material around it to secure electron equilibrium. The chamber had been calibrated by the National Bureau of Standards. The conversion factor from roentgens to rads was 0.96.

A total of 80 pigeons was used: 60 received 2,500 rads, and 20 served as unirradiated controls. The animals were separated into 4 experimental groups.

Group 1 consisted of 20 irradiated birds, 10 of which were treated in July, 1958 (Series A), while the remaining were treated in September (Series B). These birds were each given a single intravenous injection of a saline suspension of fresh homologous bone marrow within 3 hours after irradiation. The volume was adjusted so that each bird received 2 ml. of a suspension containing approximately 250×10^6 nucleated cells.

The homologous bone marrow was obtained from the femora and tibiotarsi of 10 backcross hybrids which were derived from mating to *livia* of hybrids and selected backcross hybrids from crosses between *C. guinea* and *livia*. These birds have been sufficiently backcrossed to be considered genetically homologous with *livia*, except for the genetic material producing one or both of two erythrocytic antigens, B and C, derived from *guinea*.[6, 9] Such markers on the donor's erythrocytes allowed detection in the host's circulation of blood cells derived from the bone marrow transplant because neither of these factors, B or C, is present in *livia*. In addition, a separation of the erythrocytes of the host from those of the donors was possible because the erythrocytes of each host gave positive agglutination with saline extracts of lima beans Sieva (*Phaseolus limensis*), while those of each donor were not agglutinated by this reagent.[12] Each donor supplied bone marrow for two irradiated hosts, and at no time were marrow samples from two donors pooled.

Group 2 consisted of 20 irradiated birds (10 in Series A and 10 in Series B), each of which received 2 ml. of a saline suspension of fresh heterologous bone marrow within 3 hours after irradiation. Each host received approximately 250×10^6 nucleated cells, which were obtained from femora and tibiotarsi of a single ring dove (*Streptopelia risoria*).

The 20 irradiated pigeons in Group 3 (10 each in Series A and B) received a single intravenous injection of 2 ml. of the sterile saline. Group 4 consisted of 20 unirradiated animals, 10 each in Series A and B, which were injected as in Group 3.

Two birds from each of Groups 1–4 were caged together following the irradiation

in an effort to reduce the influence of differential environmental factors to a minimum.

Blood samples were drawn from the wing vein of each animal on days 0, 4, 6, 8, 10, 12, 14, 16, 20, 24, 28, and 30 or until the animal died. Hematocrit readings on these samples determined the approximate volumes of red and white cells. The samples from Group 1-A and B were tested with antisera specific for the B and C antigenic factors, to reveal the presence or absence of the donor's erythrocytes, and with Sieva extract to detect the host's red cells.

The samples from Group 2-A and B were tested with rabbit anti-*livia* serum (absorbed with cells of *risoria*, rendering it specific for *livia* erythrocytes in contrast to those of *risoria*) to detect pigeon (host) red cells. These samples were also tested for the presence or absence of donor's red cells with an antiserum (rabbit anti-*risoria* serum absorbed with cells of *livia*) which was reactive with *risoria* erythrocytes, not with those of *livia*.

In Series B, serum samples were obtained at each bleeding from the individuals in Groups 1–4. These sera were tested for possible hemagglutinins against a standard panel consisting of erythrocyte suspensions from *livia*, *risoria*, and backcross hybrids carrying the *guinea* factors B and C. In addition, each serum was tested against cells of the bone marrow donor and its own erythrocytes obtained on day 0.

All birds were weighed at the intervals above for blood sampling and those which died were weighed as soon as possible after death.

TABLE 1

SEROLOGICAL DATA REPRESENTATIVE OF RESULTS OBTAINED IN ERYTHROCYTE REPOPULATION STUDIES OF IRRADIATED PIGEONS INJECTED WITH HOMOLOGOUS BONE MARROW (GROUP 1)*

Bird Number	Blood Type of Donor	Reagents Used to Test RBC	Agglutination Reactions of Erythrocytes† (Days Postirradiation) 0	4	6	8	10	12	14	20	28	70	Survival Time (Days)
3095 20 (Ser. B)	B	Sieva	+++	+++	+++	+++							9
		Anti-B	0	0	0	0							
		Anti-C	0	0	0	0							
3090 40 (Ser. A)	B & C	Sieva	+++	+++	+++	+++	++	++	++	++	0		30
		Anti-B	0	0	0	0	±	++	+++	+++	+++		
		Anti-C	0	0	0	0	++	++	+++	+++	+++		
3095 13 (Ser. B)	C	Sieva	+++	+++	+++	+++	++	++	++	+			21
		Anti-B	0	0	0	0	0	0	0	0			
		Anti-C	0	0	0	±	+	++	+++	+++			
3095 29 (Ser. B)	B	Sieva	+++	+++	+++	+++	nt	++	++	++	0	0	78
		Anti-B	0	0	0	0	nt	++	++	++	+++	+++	
		Anti-C	0	0	0	0	nt	0	0	0	0	0	

† Degrees of agglutination were as follows: complete (++++); approx. 75 per cent (+++); approx. 50 per cent (++); approx. 25 per cent (+); detectable but less than 20 per cent (±); and not detectable (0). "nt" = sample not tested.

* Complete data on all the individuals in these studies may be obtained upon request.

TABLE 2*

SEROLOGICAL DATA FOR 3 OF IRRADIATED PIGEONS INJECTED WITH HETEROLOGOUS BONE MARROW (GROUP 2), SHOWING EACH OF DIFFERENT TYPES OF RESULTS OBTAINED

BIRD NUMBER	TYPES OF OBSERVATIONS	TIME BLOOD SAMPLES WERE OBTAINED (DAYS POSTIRRADIATION) 0	4	6	8	10	12	SURVIVAL TIME (DAYS)
3090 27 (Ser. A)	Degree of "in vivo agglutination"	0	?†	+ + +				6
	Hemagglutinin activity of serum with RBC's of: *livia*	nt	nt	nt				
	risoria	nt	nt	nt				
3095 38 (Ser. B)	Degree of "in vivo agglutination"	0	+ +					5
	Hemagglutinin activity of serum with RBC's of: *livia*	0	+ +					
	risoria	0	0					
3095 37 (Ser. B)	Degree of "in vivo agglutination"	0	+	+ + +	+ +	0	0	13
	Hemagglutinin activity of serum with RBC's of: *livia*	0	±	+ + +	+ + + +	+ + + +	+ + + +	
	risoria	0	0	0	0	0	0	
	RBC activity with: Anti-*livia*	+ + +	+ + +	+ + +	+ + + ‡	0	0	
	Anti-*risoria*	0	+	+ + +	+ + +	+ + +	+ + +	
	Isotonic saline	0	+ +	+ + +	+ +	0	0	

* Symbols for degrees of agglutination same as used in Table 1. Complete data on all the individuals in these studies may be obtained upon request.

† "?" indicates sample not examined for "in vivo agglutination"; this was the bird that led to the discovery of the phenomenon on day 6.

‡ Agglutination in these samples is due to "in vivo agglutination," though it appears that specific antisera increase the degree of agglutination.

Autopsies were not conducted, since most of the birds were not discovered until several hours (3–8) after death.

RESULTS

GROUP 1. The results for this group of 20 animals, which received homologous bone marrow after being subjected to 2,500 rads total-body irradiation, appeared to fall into two distinct categories:

Birds Which Died within 10 Days after Irradiation.—Eleven of the 20 birds receiving homologous bone marrow died between days 5 and 10. Of these, 7 were from Series A and 4 from Series B. These pigeons showed a rapid decrease in packed cell volume of both erythrocytes and white cells. This was accompanied by severe weight loss, and, in most cases, both nausea and diarrhea preceded death.[9] There was no indication of erythrocyte repopulation by the marrow transplant in any of these animals (for typical example see bird No. 3095.20, Table 1).

Birds Which Survived Longer than 15 Days.—Of the 3 birds in Series A which survived 10 days, 1 died on day 16, 1 on day 21, and 1 on day 30. The 6 birds of Series B which survived beyond the tenth day died subsequently on days 16, 21, 26, 29, 50, and 78. In all birds which survived more than 10 days there was definite evidence of erythrocyte repopulation by the donor marrow. That is, the red blood cells of the 9 survivors were

agglutinated by antisera specific for the donor's blood type, anti-B and/or anti-C (see birds Nos. 3090.40, 3095.13, and 3095.29 in Table 1). The earliest evidence of repopulation appeared 8 days after transplantation (see No. 3095.13, Table 1), though in some birds repopulation was not demonstrated until day 12.

In the test with host-specific agglutinin (Sieva extract), a definite reduction in activity occurred, paralleled by an increased activity with donor-specific antisera beginning with the sample obtained on day 8 (see birds Nos. 3090.40, 3095.13, and 3095.29 in Table 1). The erythrocytes of all 4 birds which survived 28 days or more were no longer agglutinated by host-specific agglutinin but were, however, completely reactive with donor-specific antisera, indicating total replacement of erythropoiesis by the transplant marrow (see birds Nos. 3090.40 and 3095.29, Table 1).

GROUP 2. The 20 animals which received 2,500 rads total-body irradiation followed by heterologous bone marrow transplantation can be separated into 2 classes. The serological data for 3 of the birds in this group which show each of the different types of results obtained are shown in Table 2.

Birds Which Died Prior to Sampling on Day 6.—Of the 12 birds which died between days 4 and 6, 4 were from Series A and 8 from Series B. Hematocrit readings on day 4 revealed that 11 of these birds had subnormal erythrocyte and white cell volumes. The other bird died prior to sampling on day 4. These 12 pigeons manifested weight loss, and in most cases symptoms of nausea and diarrhea were noted prior to death.

The blood sample obtained on day 6 from bird No. 3090.27 (Series A) appeared to be agglutinated as it dropped from the wing vein and ran down the side of the collection tube (see Table 2). Second and third samples, taken from this pigeon, 1 and 2 hours later, confirmed that the erythrocytes were definitely agglutinated. This evidence indicates that such agglutination occurred in vivo; hence the term "in vivo agglutination" would apply.

The day-4 samples from birds in Series A were not examined for "in vivo agglutination," since such a reaction had not been anticipated. All 7 birds of Series B, which died prior to the day-6 bleeding, yielded blood samples on day 4 showing "in vivo agglutination." Serum samples from these birds on day 0 and 4 (with one exception) contained no demonstrable circulating agglutinin in excess of that in the agglutinated cell mass. Bird No. 3095.38 (see Table 2), whose day-4 sample was strongly agglutinated "in vivo," yielded serum on that day which showed definite agglutination with blood samples from 9 pigeons, although it was negative with 7 dove samples. Erythrocyte suspensions of the day-0 samples of the host, pigeon number 3095.38, and the bone marrow donor, dove No. D.T. 20, were included.

Birds Which Died between Days 6 and 14.—Seven birds died during this period (4 on day 6, 1 on day 7, and 2 on day 8). All but 1 bird (which died before the phenomenon had been discovered) in Series A showed strong "in vivo agglutination" at the time of day-6 bleeding.

The sera of 2 birds in Series A contained an agglutinin which was specific for pigeon erythrocytes (host species) in contrast to dove red blood cells (donor species). "In vivo agglutination" was obvious on day 4 and appeared to be more pronounced on days 6 and 8 in one bird in Series B which died on day 8. On days 4, 6, and 8

circulating agglutinins specific for pigeon red cells (in contrast to those of doves) were present, while the day O serum samples revealed no agglutinin activity for either pigeon or dove cells.

Bird No. 3095.37 of Series B (Table 2), whose day-O sample was entirely normal, survived until day 13 and showed "in vivo agglutination" beginning with the day-4 sample. The sample on day 6 appeared to be more completely agglutinated, while the degree of "in vivo agglutination" seemed to be reduced in the day-8 sample and not detectable (even microscopically) on days 10 and 12. The day-4 sample was specifically reactive with pigeon red cells, even those obtained from this host on day 0. This host-species-specific hemagglutinin activity was stronger on day 6 and remained apparently constant for the 8-, 10-, and 12-day serum samples, in spite of the absence of any indications of "in vivo agglutination" on days 10 and 12. A test of the blood samples from this bird on days 10 and 12 revealed that the erythrocytes were completely reactive with dove-specific antisera. Tests with pigeon-specific antisera were negative, even under microscopic examination, indicating total replacement of erythropoiesis by the transplanted dove bone marrow.

GROUP 3. All 20 birds in this group died between days 5 and 17, with an average survival time of approximately 8 days. These birds manifested increasing anemia, weight loss, nausea, and diarrhea. None of the blood samples from birds in Series B showed "in vivo agglutination" or agglutinin activity with erythrocytes of *livia* or *risoria*.

GROUP 4. The 20 birds in this group showed no anemia, emaciation, nausea, or diarrhea and are still apparently normal in all respects at the time of this report. As in Group 3, neither "in vivo agglutination" nor agglutinin activity was demonstrable.

DISCUSSION

The results of this preliminary work indicate that bone marrow transplantation alone may not be a satisfactory protective treatment for acute lethal irradiation damage in the species *C. livia*. In fact, those birds treated with heterologous marrow died slightly earlier, on the average, than those lethally irradiated birds receiving only saline. Though 6 of 20 of the irradiated pigeons which were injected with homologous marrow survived considerably longer than did irradiated controls, none has approached the survival time obtained with mice.[2, 3, 5, 13]

Lack of prolonged survival in *livia* was possibly due to transplantation of inadequate amounts of bone marrow, though this is probably not the only contributing factor. Intestinal damage may prove to be more serious in *C. livia* than has been suggested for mice and other mammals.[1] The fact that a large percentage of the irradiated pigeons which received homologous transplants died within the first 7 days, showing symptoms of severe nausea accompanied by diarrhea, could be indicative of gastrointestinal damage. This, coupled with the observation that a substantial number of birds died during a period after detectable erythrocyte repopulation and increasing red cell count, does not suggest anemia as the cause of death. The symptoms associated with these deaths appear to be more closely allied with those reported for rabbits, which Porter and Murray[11] attributed to bacterial and viral infection. These investigators have demonstrated that treatment with anti-

biotics in addition to the bone marrow transplant increases survival time for lethally irradiated rabbits.

Modification of the dove bone marrow transplantation technique sufficiently to obtain therapeutic effects is complicated by the hemagglutinin produced in pigeons so treated. This host-species-specific agglutinin appears very rapidly and is sufficiently potent to result in a very high degree of "in vivo agglutination," and this, superimposed on the cessation of host erythropoiesis, may prove to be therapeutically insurmountable. Characterization of this hemagglutinin from the present data partially elucidates the mechanism involved. Since no agglutinin could be demonstrated in either the host's or the donor's circulation prior to irradiation and transplantation, it seems logical to assume that this agglutinin was either produced or released after treatment. Furthermore, the fact that such hemagglutinins were not detectable in either the saline-treated or the homologously treated birds tends to rule out the possibility of viral or bacterial hemagglutinins. For the same reason, it is highly unlikely that these agglutinins are radiation-induced. The most plausible interpretation is that these agglutinins were elaborated by the cells of the heterotransplants.

Whether this agglutinin is an antibody or a complex of antibodies elicited by antigenic components of the host remains unanswered. The present data show that the specificity of the "agglutinin" is comparable to the specificity anticipated if *risoria* were immunized with the cells of *livia*. As far as has been investigated, these hemagglutinins meet all requirements of an antibody produced by the foreign graft against an antigen or antigens of the host.

The fact that this reaction occurs very early following the transplantation and is of such magnitude is possibly only a reflection of the high metabolic rate in avian species. Here we may also be dealing with bone marrow which contains a relatively high proportion of immunologically active cells, since lymph nodes, comparable to those present in mammals, are presumably absent in avian species.

SUMMARY

1. Sexually mature pigeons (*C. livia*) were subjected to 2,500 rads total-body irradiation and subsequently injected intravenously with: (1) homologous bone marrow (backcross hybrids possessing erythrocytic antigens, B and/or C, derived from *C. guinea*); (2) heterologous bone marrow (ring dove, *S. risoria*); or (3) sterile saline (irradiated controls.)

2. The 9 out of 20 birds of the homotransplant group which survived more than 10 days showed definite evidence of erythrocyte repopulation (initial appearance of donor-derived red cells was between days 8 and 12). Total replacement of erythropoiesis by the homotransplant was demonstrable on day 28 in 4 out of 20 of the birds so treated. These birds died subsequently without any indication of chimera reversal, but prolonged survival comparable to that observed in mice was not obtained.

3. Heterotransplantation resulted in the rapid appearance (4–6 days) of host-species-specific hemagglutinins sufficiently potent to produce a high degree of "in vivo agglutination." The most plausible interpretation is that such agglutinins were elaborated by the cells of the heterotransplant.

* From the Department of Genetics (paper No. 712) and Department of Radiology, University of Wisconsin. This project was supported in part by the Research Committee of the Graduate School from funds supplied by the Wisconsin Alumni Research Foundation and in part by a research grant and C-4051(R1) to H. Vermund) (G-3884 to M. R. Irwin) grants from the Department of Health, Education, and Welfare, of the National Institutes of Health, Public Health Service.

[1] C. C. Congdon, Delta Uphoff, and Egon Lorenz, *J. Natl. Cancer Inst.*, **13,** 73–107, 1952.

[2] D. W. H. Barnes, C. E. Ford, P. L. T. Ilbery, and J. F. Loutit, *Transpl. Bull.*, **5,** 101–106, 1958.

[3] D. W. van Bekkum and O. Vos, *J. Cellular Comp. Physiol.*, **50,** 139–156, 1957.

[4] C. E. Ford, J. L. Hammerton, D. W. H. Barnes, and J. F. Loutit, *Nature*, **177,** 452–454, 1956.

[5] N. Gengozian, T. Makinodan, C. C. Congdon, and R. D. Owen, these PROCEEDINGS, **44,** 560–565, 1958.

[6] M. R. Irwin, L. J. Cole, and C. D. Gordon, *J. Exptl. Zoöl.*, **73,** 285–308, 1936.

[7] D. L. Lindsley, T. T. Odell, and F. G. Tausche, *Proc. Soc. Exptl. Biol. Med.*, **90,** 512–515, 1955.

[8] T. Makinodan, *Proc. Soc. Exptl. Biol. Med.*, **92,** 174–179, 1956.

[9] W. J. Miller and C. R. Bryan, these PROCEEDINGS, **39,** 407–412, 1953.

[10] T. T. Odell, F. G. Tausche, D. L. Lindsley, and R. D. Owen, *Ann. N.Y. Acad. Sci.*, **64,** 811–823, 1957.

[11] K. A. Porter, and J. E. Murray, *J. Natl. Cancer Inst.*, **20,** 189–206, 1958.

[12] D. H. Shaw, and J. H. Stimpfling, unpublished material.

[13] J. J. Trentin, *Proc. Soc. Exptl. Biol. Med.*, **96,** 139–144, 1957.

ON THE NATURE OF THE AGING PROCESS

BY LEO SZILARD*

ENRICO FERMI INSTITUTE OF NUCLEAR STUDIES, UNIVERSITY OF CHICAGO

Communicated by Theodore Shedlovsky, November 24, 1958

Introduction.—This paper represents an attempt to describe a hypothetical biological process that could conceivably account for the phenomenon of aging. Aging manifests itself in much the same general manner in all mammals, and we are in a position to learn enough about the aging of mammals to be able to test the validity of a theory that leads to predictions of a quantitative kind—as does the theory here presented.

We know that a gene can be responsible for the synthesis of a specific protein molecule, which in many cases has a known enzymatic activity. When we speak later of a mutant, or "incompetent," form of a gene, we mean an altered form of the gene, which cannot synthesize the specific protein molecule in its chemically active form.

Our theory assumes that the elementary step in the process of aging is an "aging hit," which "destroys" a chromosome of the somatic cell, in the sense that it renders all genes carried by that chromosome inactive. The "hit" need not destroy the chromosome in a physical sense. (See note 1 added in proof.)

We assume that the "aging hits" are random events and that the probability that a chromosome of a somatic cell suffers such a "hit" per unit time remains constant throughout life. We further assume that the rate at which chromosomes of

a somatic cell suffer such "hits" is a characteristic of the species and does not vary appreciably from individual to individual.

As a result of an aging process of this nature, the number of the somatic cells of an individual organism which have "survived" up to a given age (in the sense of remaining able to fulfil their function in the organism) decreases with age. On the basis of our assumptions, spelled out below, the "surviving" fraction of the somatic cells decreases with age at an accelerating rate.

Our theory postulates that when f, the surviving fraction of the somatic cells of an individual, approaches a certain critical value, f^*, then the probability that that individual may die within a period of one year will come close to 1. On this basis, the theory establishes a relationship between the surviving fraction of the somatic cells and the age of death of the individual.

Because the young mammalian organism may be assumed to have a large functional reserve, we shall assume that the surviving fraction of the somatic cells of an individual may fall substantially before the organism loses its capacity to live, perhaps to a value somewhere between $^1/_3$ and $^1/_{12}$.

The precise meaning of the term "critical value," f^*, will shift as we go from the crudest form of the theory, which we shall discuss first, to a less crude form of the theory, which we shall discuss thereafter. In the crudest form of the theory, we shall assume that an adult does not die of natural causes until the surviving fraction of his somatic cells comes very close to the critical fraction f^* and that he dies at the critical age, i.e., within the year in which this surviving fraction reaches the critical fraction f^*. Thus, in its crudest form, the theory postulates that the age at death is uniquely determined by the genetic makeup of the individual.

Clearly, this cannot be strictly true, for, if it were true, identical twins would die within one year of each other. In fact, the mean difference between the ages at death of female identical twins can be estimated to be about 3.5 years. The discrepancy arises from the failure of the crude theory to take into account that in a cohort of identical individuals the number of deaths per year may be expected to rise as a continuous function with advancing age and that an appreciable number of deaths may be expected to occur at ages lower than the "critical age."

If not otherwise stated, our discussion here relates to man and, in particular, to the female of the species. In the case of man, the somatic cells of the female contain $m = 23$ pairs of homologous chromosomes. There may be in man perhaps 15,000 genes. There may be a larger number of specific DNA molecules which are inherited from generation to generation, but we designate as "genes" here only those DNA molecules which would handicap the individual if present in a mutant form. An individual who is a heterozygote for a mutant gene might not necessarily be handicapped under the conditions prevailing at present in the United States, where essentially no adult dies for lack of food or shelter and no adult has a reduced propensity to procreate because of his inability to provide food or shelter for his offspring. But such a heterozygote would have been handicapped (according to our definition of the term "gene") under conditions which were prevalent in the past—up to recent times. The present abundance of mutant forms of genes in the population may not correspond to the steady state under present conditions.

We may assume that the "genes" somehow affect differentiation and morphogenesis during the embryonic development of the individual and that a mutant form

of a gene may cause, with a certain probability—appreciable even in the heterozygote—a developmental abnormality of the individual.

We assume that among the 15,000 genes, there are perhaps 3,000 genes which are important for the functioning of the somatic cells of the adult. We shall call these genes "vegetative genes," and a mutant form of such a gene we shall designate as a "fault." Of the remainder of the genes, we shall assume that they are irrelevant for the functioning of the somatic cells of the adult organism.

We postulate that, in the course of aging, a somatic cell remains functional as long as, out of each pair of homologous vegetative genes, at least one of the two genes is competent and active and that the cell ceases to be functional when both genes are out of action. Accordingly when a chromosome suffers an aging hit, the cell will cease to be functional if the homologous chromosome has either previously suffered an aging hit or if it carries a fault.

According to the views here adopted, the main reason why some adults live shorter lives and others live longer is the difference in the number of faults they have inherited. If we assume that faults are distributed in the population at random, then we can compute the distribution of the faults, from the mean value of faults per person (which we shall designate by n). From the observed distribution of the ages at death, between seventy and ninety years of age, we shall be led to conclude that we have $n > 2$. For $n = 2$ we would obtain from the crude theory for the critical surviving fraction of the somatic cells $f^* \backsimeq 1/4$. For $n = 4$ we would obtain $f^* \backsimeq 1/12$. On this basis we shall be led to conclude that we have $n < 4$.

We shall, for the purposes of our discussion, adopt, as a reasonable value, $n = 2.5$, and then we obtain $f^* \backsimeq 1/6$, which would seem to be a reasonable value.

The "Surviving" Fraction of the Somatic Cells.—We shall now proceed to compute the "surviving" fraction of the somatic cells of a female who has inherited r faults, as a function of her age.

We designate by ξ the average number of "aging hits" that have been suffered by a chromosome of a somatic cell, and we may write

$$\xi = \frac{1}{2m} \frac{\text{age}}{\tau}, \tag{1}$$

where τ is the average time interval between two subsequent aging hits suffered *in toto* by the m pairs of homologous chromosomes contained in a somatic cell. We may call this average time interval τ the "basic time interval of the aging process."

Let us now consider a female who has inherited r faults. If none of the pairs of homologous chromosomes contain more than one fault—a condition likely to be fulfilled if r is small compared to m—then we may write for the "surviving" fraction of her somatic cells at a given age

$$f = [1 - (1 - e^{-\xi})^2]^{m-r} \cdot e^{-r\xi} \tag{2}$$

or

$$\ln f = (m - r) \ln [1 - (1 - e^{-\xi})^2] - r\xi. \tag{3}$$

For $\xi \ll 1$ we may write, from equation (3), neglecting $m\xi^4$ and $r\xi^3$, etc.,

$$\ln \frac{1}{f} = m(\xi^2 - \xi^3) + r(\xi - \xi^2). \tag{4}$$

Writing

$$\rho = \frac{r}{2m}, \tag{5}$$

we may write

$$\ln \frac{1}{f} = m[\xi + \rho] \cdot [1 - (\xi + \rho)], \tag{6}$$

provided $r \ll x \ll 2m$ (i.e., $\rho \ll \xi - 1$)

$$\ln \frac{1}{f} = m\eta^2(1 - \eta), \tag{7}$$

where $\eta = \xi + \rho$. In place of equation (7) we may write, in our approximation,

$$f = [1 - (1 - e^{-\eta})^2]^m \tag{8}$$

We may also write inversely

$$\eta = \ln \frac{1}{1 - \sqrt{1 - f^{1/m}}}, \tag{9}$$

or, expanding,

$$\eta = \sqrt{\frac{1}{m} \ln \frac{1}{f}} + \frac{1}{2m} \ln \frac{1}{f}. \tag{10}$$

According to the assumption of the crude theory, f, the surviving fraction of the somatic cells, reaches the critical value f^* at the age of death, which we designate by t_r. Further, we designate by x_r the average number of aging hits suffered *in toto*, up to the age of death, by the m pairs of homologous chromosomes of the somatic cells. Thus we have, at the age of death,

$$x_r = \frac{t_r}{\tau} = 2m\xi \tag{11}$$

and

$$\eta = \frac{x_r + r}{2m}. \tag{12}$$

Accordingly, we may write at the age of death, where we have $f = f^*$, from equation (8),

$$\ln \frac{1}{f^*} = \frac{(x_r + r)^2}{4m} \left(1 - \frac{x_r + r}{2m}\right). \tag{13}$$

Similarly, we may write at the age of death, from equation (10),

$$x_r + r = \sqrt{4m \ln \frac{1}{f^*}} + \ln \frac{1}{f^*} \tag{14}$$

or

$$\frac{t_r}{\tau} + r = \sqrt{4m \ln \frac{1}{f^*}} + \ln \frac{1}{f^*}. \tag{15}$$

For the genetically perfect female, for whom we have $r = 0$, we shall designate the age at death by t_0. We shall call t_0 the "life-span" of the species.

From equation (15) we may write for, the life-span, t_0,

$$\frac{t_r}{\tau} + r = \frac{t_0}{\tau} \tag{16}$$

or

$$t_r = t_0 - \tau r \tag{17}$$

or

$$r = \frac{t_0 - t_r}{\tau}. \tag{18}$$

As may be seen from equation (17), the addition of one fault to the genetic makeup of an individual shortens the life of that individual by $\Delta t = \tau$, so that we may write

$$\Delta t \text{ per fault } = \tau. \tag{19}$$

This expresses one of the basic results of our theory. According to equation (19), an individual whose genetic makeup contains one fault more than another individual has a life-expectancy which is shorter by τ, the basic time interval of the aging process. This holds true within the crude theory for individuals who have inherited a small number of faults.

Concerning the life-span, t_0, we may write, from equations (11), (13), and (16),

$$\ln \frac{1}{f^*} = \frac{1}{4m} \left(\frac{t_0}{\tau}\right)^2 \left(1 - \frac{1}{2m} \frac{t_0}{\tau}\right), \tag{20}$$

and, from equations (15) and (16), we may write

$$\frac{t_0}{\tau} = \sqrt{4m \ln \frac{1}{f^*}} + \ln \frac{1}{f^*}. \tag{21}$$

The Distribution of the Ages at Death.—The above equations hold within the framework of the crude form of the theory. In this form of the theory, members of one cohort would die only in certain years—at the critical ages, t_r—and thus the years in which death occurs within one cohort would be separated from each other by time intervals of τ years; no deaths would occur in the intervening years.

Further, if the distribution of the faults in the population is random, then the number of deaths, P_r, occurring at each age, is given by the Poisson distribution:

$$P_r = \frac{n^r e^{-n}}{r!} \tag{22}$$

where, according to equation (18), we have $r = \frac{t_0 - t_r}{\tau}$ and where n stands for the average number of faults per individual.

The distribution of the ages at death in the population is actually a continuous function of the age. Even though the probability that an individual may die within a year may increase rather steeply as the surviving fraction of his somatic cells approaches the critical value f^*, genetically identical individuals do not all die at the same age. The observed mean age difference at death of identical twins may be regarded as a measure of the scattering of the ages at death, which is left out of account by the crude form of the theory and to which we shall refer as the "non-genetic scattering."

For the time being, we shall continue to leave this non-genetic scattering out of account; yet, for the sake of convenience, we shall henceforth describe the distribution of the ages at death by $P(r)$, a continuous function of r, in place of the discontinuous "Poisson" values, P_r. For $P(r)$ we may write

$$P(r) = \frac{n^r}{\Gamma_{(r+1)}} e^{-n} \tag{23}$$

where Γ represents the gamma function (which for integral values of $r + 1$ assumes the values of $r!$) and where we have

$$r = \frac{t_0 - t}{\tau}. \tag{24}$$

For the number of deaths occurring within a cohort per unit time, we may then write, according to our theory,

$$d(\text{theor.}) = -\frac{dr}{dt} \frac{n^r}{\Gamma_{(r+1)}} e^{-n}. \tag{25}$$

From equation (24) we obtain

$$-\frac{dr}{dt} = \frac{1}{\tau}. \tag{26}$$

Thus we may write, from equation (25),

$$d(\text{theor.}) = \frac{1}{\tau} \frac{n^r}{\Gamma_{(r+1)}} e^{-n} \text{ per year}, \tag{27}$$

where r is given by equation (24) and where τ is expressed in years.

The approximation used throughout this paper holds for small values of r, which correspond to high ages at death. We may say that at high ages of death the distribution of ages at death in the population is represented by a reversed Poisson distribution (27), where small values of r correspond to high ages at death.

Lower Limit for n.—We shall now proceed to compare the distribution of the ages at death, as given by our formula (27), with the actually observed distribution of the ages at death, as given by the U.S. Life Tables, based on the 1949–50 Census.

For the purposes of this comparison, we shall use Table 6 for white females, which lists the number of deaths per year, in yearly intervals, as a function of age. According to this table, the maximal number of deaths occurs between the eightieth and eighty-first year; the corresponding maximal number of deaths per year is 0.0344 per person.

The distribution of the ages at death is not symmetrical around the age at the maximum, $t^* = 80.5$ years of age; the number of deaths per year fall faster toward higher ages than toward lower ages. Thus the table lists, for the number of deaths per year, 0.0230 per person between the ages seventy and seventy-one and 0.0179 per person between the ages of ninety and ninety-one.

We may derive from this table a "normalized" distribution of the ages at death by forming R(obs.), the ratio of the number of deaths per year and the maximal number of deaths per year, 0.0344. Thus we obtain R(obs.) = 0.667 at 70.5 years of age; R(obs.) = 1 at 80.5 years of age; R(obs.) = 0.520 at 90.5 years of age.

We may similarly obtain from the number of deaths per year, given as a function of age by the theory, a "normalized" distribution of the ages at death, by forming R(theor.), the ratio of the number of deaths per year given by equation (27) and the maximal number of deaths per year given by $d(\text{theor.})_{\text{max.}}$:

$$d(\text{theor.})_{\text{max.}} = \frac{1}{\tau} e^{-n} \left\{ \frac{n^r}{\Gamma_{(r+1)}} \right\}_{\text{max.}}. \tag{28}$$

If we designate by r^* the value of r for which this expression becomes maximal, we may write, for $n \geqq 2$,

$$r^* \backsimeq n - 0.5. \tag{29}$$

Accordingly, we may write

$$R(\text{theor.}) = \frac{d(\text{theor.})}{d(\text{theor.})_{\text{max.}}} = \frac{n^r}{n^{(n+0.5)}} \frac{\Gamma_{(n+0.5)}}{\Gamma_{(r+1)}}. \tag{30}$$

We may now ask for what value of n would the normalized Poisson distribution, R(theor.), fit R(obs.), both above and below 80.5 years of age so that we have for a suitably chosen value of Δr, for $r = r^* + \Delta r$, R(theor.) = 0.667 (the value of R(obs.) at 70.5 years of age) and that we also have, for $r = r^* - \Delta r$, R(theor.) = 0.520 (the value of R(obs.) at 90.5 years of age).

It turns out that such a fit is possible only for a value of n which is very close to $n = 2$. For the corresponding value of Δr we obtain $\Delta r = 1.4$. For the corresponding value of τ we may write

$$\tau = \frac{10}{\Delta r} \text{ years}. \tag{31}$$

For $n = 2$ and with $\Delta r = 1.4$, we obtain $\tau = 7.15$ years.

For values of n which are substantially larger than 2, it is not possible to fit the normalized Poisson distribution R(theor.) to R(obs.) in this manner. If R(theor.) is made equal to 0.520 (the value of R(obs.) at 90.5 years of age) for $r = r^* - \Delta r$, then, for $r = r^* + \Delta r$, we have R(theor.) < 0.667 (the value of R(obs.) at 70.5 years of age).

Because there is reason to believe that, below 80.5 years of age, the crude theory gives too low values for R(theor.), we cannot exclude the possibility that we have $n > 2$. Therefore, from the fact that R(theor.) derived from the crude theory fits R(obs.) for $n = 2$ between the ages of 70.5 and 90.5, we may not conclude that we actually have $n = 2$, and we may only conclude that we have

$$n \geqq 2. \tag{32}$$

Approximation of the Poisson Distribution by a Gaussian Relationship between τ and n.—R (theor.) given by equation (30) goes over into a Gaussian for $n \gg 1$. For a Gaussian, the value of R(theor.) = 0.667 (the value of R(obs.) at 70.5 years of age) corresponds to a distance from the maximum of 0.9 σ, where σ is the standard deviation of the Gaussian. Similarly, R(theor.) = 0.520 (the value of R(obs.) at 90.5 years of age) corresponds to a distance from the maximum of 1.14σ. Thus the time interval of 20 years around the maximum corresponds to 2.04σ, and hence we have

$$\sigma = \frac{20}{2.04} \text{ years} = 98 \text{ years} \tag{33}$$

Because the variance of a Poisson distribution is given by its mean, n, we may write (for $n \gg 1$)

$$\sigma = \tau\sqrt{n} \tag{34}$$

or

$$\tau = \frac{\sigma}{\sqrt{n}}, \tag{35}$$

and thus we obtain $\tau = \dfrac{9.8}{\sqrt{n}}$ years.

While equation (34) holds, strictly speaking, only for large values of n, the error is small even for $n = 2$.

For $n = 2$ from equation (35) we obtain $\tau = 6.82$ in place of the previously given value of $\tau = 7.15$.

For $n = 2.5$ from (35) we obtain $\tau = 6.2$ years in place of the "correct" value of $\tau = 6.3$ years, which we find by fitting the "normalized" Poisson distribution, as well as possible, to R(obs.).

Thus, for most of our purposes we may use equation (35) for values of n for $n \geqq 2$.

From equation (35) we obtain

$$n\tau = \sigma\sqrt{n}, \tag{36}$$

where $n\tau$ represents the average life-shortening caused by the "load of faults," n. Since the value of σ is empirically fixed, the higher we assume n to be, the higher is the life-shortening effect which we must attribute to it. In this sense, the life-shortening effect of the mutation load increases with $\sqrt{n}$.

Correction of τ for the Non-genetic Scattering of the Ages at Death.—Because the non-genetic scattering of the ages at death has so far not been taken into account by us, the observed distribution of the ages at death may be expected to be actually somewhat broader, and, accordingly, the actual value of τ may be expected to be somewhat lower than the values given above.

The mean age difference at death between female identical twins has been reported by Franz J. Kallman to be about 2.6 years for twins dying above the age of 60. From this value we may estimate, on the basis of the Life Tables, the mean age difference at death of female identical twins who die as adults above the age

of 40 to be 3.4 years. If the distribution of the ages at death of genetically identical individuals resembled a Gaussian, then the variance of the distribution of the ages at death in the population would be equal to the sum of the variance of this Gaussian and that of our theoretical distribution of the ages at death. By making such an assumption, for the purposes of this computation, we may then correct the values of τ, given above, as follows:

From the fact that the mean age difference at death of female identical twins may be taken to be about 3.4 years, it follows that the standard deviation of the distribution of their ages at death is about 3 years. Using this value, we find that the nongenetic scattering increases the variance of the distribution of the ages at death by a factor of about 1.1 and, accordingly, the previously given values of τ must be reduced by 5 per cent.

Thus we may now write, for the corrected values of τ, for $n = 2$, $\tau = 6.8$ years; and for $n = 2.5$, $\tau = 6$ years.

We may also write on this basis—within the limits of the approximation—for $n > 2$,

$$\tau = \frac{9.3}{\sqrt{n}} \text{ years.} \tag{37}$$

Substituting this value of τ in equation (16), we obtain

$$\frac{t_0}{\tau} = \frac{t_r}{9.3}\sqrt{n} + r. \tag{38}$$

The Value of the Critical Surviving Fraction of the Somatic Cells f^*—Upper Limit for n.—In order to compute the critical surviving fraction of the somatic cells, f^*, we shall now make use of the fact that (for white females) the maximal number of deaths per year occurs at 80.5 years of age. Our theory demands (29) that the maximal number of deaths per year should occur for individuals for whom we have $r = n - 0.5$. Accordingly, we may write $t_r = 80.5$ and $r = n - 0.5$. We thus obtain, from equation (38),

$$\frac{t_0}{\tau} = \frac{80.5}{9.3}\sqrt{n} + n - 0.5 \tag{39}$$

and, from equation (20),

$$\ln \frac{1}{f^*} = \frac{1}{4m}\left(\frac{80.5}{9.3}\sqrt{n} + n - 0.5\right)^2\left[1 - \frac{1}{2m}\left(\frac{80.5}{9.3}\sqrt{n} + n - 0.5\right)\right]. \tag{40}$$

From this equation we may now compute for a given value of n the corresponding value of f^*. Thus we obtain for $n = 2$, $f^* \backsimeq 1/4$; for $n = 2.5$, $f^* \backsimeq 1/6$; for $n = 4$, $f^* \backsimeq 1/12$.

On this basis we may then write, by assuming $f^* > 1/12$,

$$n < 4. \tag{41}$$

A value of $f^* \backsimeq 1/6$ would seem to be rather reasonable and, therefore, we shall adopt, as a reasonable value for n, the value of

$$n \backsimeq 2.5, \tag{42}$$

and, as a reasonable value for τ, the corresponding value of

$$\tau = 6 \text{ years.} \tag{43}$$

We shall in the remainder of the paper base all our discussions on these values of n and τ. We cannot exclude, of course, the possibility that n might be somewhat larger and that τ might be somewhat lower.

The "Physiological Age."—The general physiological age may be defined for a given population on the basis of its age-specific death rate; according to our theory, it may be defined as the age of the genetically perfect female who has the same surviving fraction of the somatic cells, f. Accordingly, we may say that two females, whose genetic makeup differs by Δ faults, differ from each other in their physiological age by $r\Delta$ years at sufficiently high ages as demanded by the approximation used.

Changing the Load of Faults.—If, as a result of living under "modern" conditions, our load of faults should, in time, be doubled, then the average adult woman would live $n\tau$ years less than she does today.

For $n = 2.5$ we have $n\tau = 15$ years. Thus the physiological age of the average female at 65 would be the same as that of the average 80-year-old woman today.

If we were to assume that $n > 2.5$, then $n\tau$ would amount to more than 15 years because $n\tau$ increases, according to equation (36), with $\sqrt{n}$.

A doubling of our load of faults might conceivably occur, in time, through the exposure of the population, generation after generation, to ionizing radiation, in an intensity that doubles the mutation rate.

Such an increase in our load of faults might perhaps occur also as a result of the current practice of controlling the family size. As spelled out below this practice might conceivably eliminate one of the selection pressures which have tended to keep our load of faults low.

We may, on this occasion, also ask how much advantage the genetically perfect (faultless) female would have over the average female of today. Assuming $n = 2.5$, we may say, on the basis of considerations similar to those presented above, that the genetically perfect female would at 50 years of age have the same physiological age as the average female of 35 today. Her most probable age at death would be 92 instead of 80. If n were larger than 2.5, the advantage of the genetically perfect female would be greater.

Life-Shortening Effect of Ionizing Radiation on the Adult Offspring of the Exposed Population.—Experiments of W. L. Russell have shown that the offspring of mice which have been exposed to a dose of fast neutrons have a reduced life-expectancy. This has generally been interpreted to mean that exposure of the parents to ionizing radiation induces mutations in the germ cells of the gonads and thus "reduces the viability" of the offspring.

From the point of view of our theory, however, we have to distinguish between that reduction of the life-expectancy of the offspring which is due to an increased mortality of young animals and that reduction which is due to a decrease in the life-expectancy of the adults. All the mutations induced by ionizing radiation may contribute to the former, but only the "faults" contribute to the latter.

In the case of man, at least, it should be possible to make a fairly clean separation between these two categories of life-shortening. In the case of man the U.S. Life

Tables show that the number of deaths per year falls, from an initial high value in the first year of life, to about 40 per 100,000 per year at the age of 10. Moreover, of the few deaths occurring at 10 years of age, a substantial fraction is due to accidents. Thus we are led to believe that, in the heterozygous individual, mutant (incompetent) forms of genes may cause the death of the embryo, or of the infant below 10 years of age, while they do not cause death with an appreciable probability after the tenth year of age, unless they represent "faults." Faults increase the age-specific death rate above 10 years of age only in conjuction with aging hits, and they increase it appreciably only above 40.

If we observe the life-shortening of the adult animal in the offspring of an irradiated population, resulting from the induction of faults by ionizing radiation in one species, we may be able to predict, on the basis of our theory, the life-shortening for another species. This may be seen as follows:

We obtain from equations (19) and (21) for the relationship between the life-shortening, Δt per fault and for the life-span of the species, t_0,

$$\frac{\Delta t \text{ per fault}}{t_0} = \frac{\tau}{t_0} = \left(\sqrt{4m \ln \frac{1}{f^*}} + \ln \frac{1}{f^*} \cdot \right)^{-1} \tag{44}$$

The right-hand side of the equation contains only the chromosome number m and the critical value f^*. Therefore, if two species of mammals have the same value of f^* and the same chromosome number m, their life-shortening per fault caused by exposure of their parents to ionizing radiation amounts to the same fraction of their life-span. We may call the ratio the "specific life-shortening" of a fault.

If the two species of mammals may be assumed to have also the same number of vegetative genes and if the sensitivity of their genes to the ionizing radiation employed may be assumed to be the same, then the number of faults produced by a given dose of radiation will be the same for the two species. Thus, according to equation (44), the radiation exposure will shorten the lives of the two species by the same fraction of the life-span.

If m, the number of their chromosome pairs, is different for the two species, then the "specific life-shortening" will be larger for the species which has the smaller chromosome number. According to (44), the specific life-shortening increases about inversely with $\sqrt{m}$. The number of chromosome pairs is 11 for the Chinese hamster and 39 for the dog. Therefore, according to (44), the specific life-shortening per fault induced may be expected for the Chinese hamster to be higher than for the dog by a factor of about 2.

The mouse has $m = 20$ pairs of homologous chromosomes and we may therefore estimate the life-shortening of man from the life-shortening of the mouse and vice versa, by postulating that the life-shortening per rep in man and in the mouse amount to about the same fraction of their life-span. (Some authors believe that man is about twice as sensitive to X-rays as the mouse and, if they are correct, then our estimated value for the life-shortening of man would be low by a factor of about two.)

W. L. Russell found that an X-ray dose of 300, as well as a dose of 600 rep, induces 25×10^{-8} gene mutations per rep per locus in the spermatogonia of mice. As-

suming 15,000 loci, we may conclude that an X-ray dose of 667 rep would induce 5 mutations in the diploid offspring of exposed mice. If we assume that one-fifth of the genes are vegetative genes, then 667 rep of X-rays induce on the average one fault in the offspring. Assuming $\tau = 6$ years for man, we may thus expect in man a life-shortening of the adult offspring of $\delta^* = 3.3$ days per rep.

The number of faults induced in the offspring per rep depends on the nature of the ionizing radiation, and may be assumed higher for fast neutrons than for X-rays. Also, in the case of X-rays, the number of faults induced might conceivably depend not only on the total dose but also on the dose rate and be lower for lower dose rates.

The actual value of δ^* ought to be determined experimentally, for the different kinds of ionizing radiations which are of interest, by direct observation of the life-shortening of the adult offspring. Experimental data so far available are inadequate.

If the average mutation rate per gene per generation is 1/60,000 and if we assume for N_1, the number of vegetative genes, $N_1 = 3000$, and for N_t the total number of genes, $N_t = 15{,}000$ then we obtain for μ_1, the spontaneous mutation rate of the haploid set of vegetative genes, $\mu_1 = 0.05$, and for μ_2, the total spontaneous mutation rate of all genes, $\mu_2 = 0.25$. We shall use these values for the purposes of our discussion below.

The Life-Shortening Effect of Ionizing Radiation on the Exposed Population.—We may expect ionizing radiation to produce gene mutations in the chromosomes of the somatic cells of an exposed individual and we shall assume that the sensitivity of the genes of the somatic cells is about the same as that of the genes in the spermatogonia and the oogonia. Because a certain fraction of these mutations, perhaps one-fifth of them, affect vegetative genes, faults are induced in the chromosomes of the somatic cells of the exposed individual. It can be shown that an exposure to ionizing radiation which induces on the average one fault per somatic cell must reduce—on this score along—the life-expectancy of the exposed individual by about τ years. If exposure to ionizing radiation had no other life-shortening effect, the life-shortening, δ, of the individual in the exposed population would be equal to δ^*, the life-shortening of the adult offspring of the exposed population (see note 2 added in proof).

Maternal Selection Pressure against Faults?—It is conceivable that a woman who carries a particular fault in her genetic makeup ceases to be capable of bearing children $\tau = 6$ years earlier, on the average, than her counterpart who lacks that particular fault. This is what one would expect on the assumption that the termination of a woman's reproductive period is determined by her physiological age—if all other factors are equal. If this assumption is correct, then a powerful selection has operated in the past that has tended to keep the load of faults low.

In the past, infant mortality was high, the birth rate was high, and women kept on having children until the end of their childbearing period. Clearly, the maternal selection mentioned above is switched off when women have two or three children between the ages of 18 and 25 and then avoid having further children.

If such a "maternal" selection was the predominating selection of the past, then we may expect that, when this selection ceases to operate, our load of faults may double, in time. As discussed before, senescence would then set in about 15 years earlier.

However, even if we assume the worst in this respect, our load of faults still cannot increase by more than $2\mu_1 = 0.1$ per generation. This means that—at the worst—it would take 25 generations for our load of faults to double.

The effect of the "maternal" selection here discussed might be estimated as follows:

Let us single out one vegetative gene. If a woman carries, as a heterozygote, this particular gene in a mutant form, her physiological age is $\tau = 6$ years higher than that of another women who does not carry this particular "fault" but who is otherwise identical in her genetic makeup. We now assume that the "physiological age" sets the termination of the reproductive period, and we take for the "most probable duration" of the reproductive period 30 years. Thus the fault singled out "most probably" shortens the reproductive period by one-fifth of its length. The fertility of younger women is higher than that of older women. Near the end of the reproductive period, the (average) time interval between two successive pregnancies might be by a factor $k > 1$ (of perhaps 2 or 3) longer, than the interval between two successive pregnancies, averaged over the whole reproductive period.

If we postulate that this "maternal" selection constitutes essentially the sole selection pressure against faults, then we may write for the mutation equilibrium of the fault, singled out:

$$2\frac{\mu_1}{N_1} = \frac{n}{2 \times 5\, kN_1}, \tag{45}$$

and, for the value of $\mu_1 = 0.05$ per generation, we thus obtain $k = n$. For $n = 2.5$ we have

$$k = 2.5. \tag{46}$$

Under "natural" conditions, for young women the average time interval between two successive pregnancies might be about 1 year; toward the end of the reproductive period it might be about 5 years; and, averaged over the whole reproductive period of about 30 years, it might be about 2 years. These values correspond to $k = 2.5$. This coincidence might, of course, turn out to be fortuitous.

Refinements of the Theory—Specialized Vegetative Genes.—It appears likely that there exist genes which are not essential for the functioning of most of the somatic cells of the adult but each of which is essential for the functioning of one particular kind of specialized somatic cell. We shall call these genes "specialized vegetative" genes, and mutant forms of such genes we shall call "specialized faults."

We shall now single out, for the purposes of our discussion, specialized cells which synthesize a gene product, a particular enzyme, for instance, that serves not the needs of these cells themselves but rather those of the organism as a whole. Such specialized cells might in some cases fulfill their function in the organism by releasing the enzyme into the circulation.

We may in general assume that the normal young person has a considerable reserve of such enzymes, and we shall specifically assume, for the purposes of this particular discussion, that the maximal output of a normal young person is higher than the need of the organism by a factor of about 6.

The maximal output of enzyme by such a specialized cell may be assumed to be

lower by a factor of $^1/_2$ in the heterozygous individual, who has inherited a mutant form of the specialized vegetative gene. Further, it may be shown that, for an individual who has inherited $n = 2.5$ faults, the surviving fraction of the somatic cells is about one-third at 54 years of age. Since, in the heterozygote, the specialized cells under discussion carry one additional fault, the surviving fraction of the specialized cells will reach one-third about 6 years earlier. Thus, at 48 years of age, the maximal output of the enzyme of the surviving specialized cells of the heterozygote will be lower, by a factor of about $^1/_6$, than for a normal young person.

On this basis we may then expect that around 50 years of age there may become manifest, in such heterozygotes, symptoms of disease due to the insufficiency of the output of one kind of specialized cell. The inheritance of diseases of this class may be expected to show a marked degree of dominance.

Speaking more generally, we may expect to see in certain heterozygotes, late in life, narrowly circumscribed degenerative phenomena which are caused by specialized faults they have inherited.

The Number of "Segments" per Chromosome.—Instead of assuming that a whole chromosome is "destroyed" in one aging hit, we might choose to assume that the elementary step in the process of aging consists in the random destruction of one-half of a chromosome. The formulas given above then remain unchanged, except that we have to write $2m$ in place of m. As one may see from equation (40), we then obtain, for the same value of n, a higher value for f^*. Thus for $n = 2.5$ we obtain $f^* \backsimeq {}^1/_3$. Apart from this, the general character of the theory remains unchanged.

However, one might ask at this point whether one could not generalize the theory, presented above, by assuming that each chromosome consists of g segments and that the elementary step in the process of aging consists in the random destruction of such segments, independently of each other. By choosing the value of g larger and larger, we might then gradually change the character of the theory and might end up with a theory which postulates that the aging process consists in a sequence of gene mutations of the chromosomes of the somatic cell.

A theory of this kind would, however, come up against difficulties, which are as follows:

As may be seen from equation (40) (where we now have to write gm in place of m), for a fixed value of f^*, n goes up roughly parallel with increasing g. A very large value of n might, however, be incompatible with the known fertility of consanguinous matings.

Further, as we increase g, we would also increase the difference of the life-expectancy of the female and the male. The male of the species has only one X chromosome, while the female has two. Let us disregard here the possibility that a substantial piece of the X chromosome might be covered, in the male, by genes contained in the Y chromosome. Let us also assume, for the sake of argument, that f^* has the same value for the male as it has for the female. On the basis of these assumptions, we may then identify the male, from the point of view of his life-expectancy, with a female who has suffered g aging hits, prior to birth. Accordingly, we may expect the adult male to live a shorter time, by $g\tau$ years, than the adult female.

Actually, according to the 1949–50 Census, the maximal number of deaths for the

white male occurs between the ages of 77 and 78, i.e., three years earlier than for the white female. This difference is three years less than what we would expect on the basis of our theory, which assumes $g = 1$ and which gives an estimate for τ of $\tau = 6$ years. This discrepancy indicates that perhaps the value of f^* is somewhat larger for the male than for the female.

Because of the possibility that this might be the case, we conceivably have $g = 2$ ($g = 2$ would mean that the elementary process of aging consists in the "destruction" of one-half of a chromosome rather than a whole chromosome).

However, there is no reason to believe that f^* may be very much larger for the male than for the female. Therefore, the observed small difference between the life-expectancy of the female and that of the male may rule out a modification of the theory that assumes $g \gg 1$.

Experimental Test of the Theory.—The most stringent experimental test of the validity of our theory is likely to come from experiments in which one observes a reduction in the life-expectancy of the adult offspring of, say, an irradiated mouse population. Experiments of this sort are needed in order to determine the value of δ^*. Experiments of this sort will also show whether among the different phenomena which generally accompany senescence, such as the graying of the hair, the loss of accommodation of the eye, etc., there are any which are determined by the general physiological age, defined on the basis of the age-specific death rate. Arrangements for experiments along these lines are now under discussion.

I am grateful to Dr. Gertrud Weiss and to Dr. N. Conant Webb, Jr., of the Medical School, University of Colorado, for detailed criticism of some of the computations involved, and to Samuel W. Greenhouse, of the National Institute of Mental Health, for critical evaluation of some of the approximations used. For carrying out auxiliary computations I am also grateful to Mrs. Dorothy Lathrop, Office of Mathematical Research, National Institutes of Health, and to Mrs. Norma French, Section on Theoretical Statistics and Mathematics, Biometrics Branch, National Institute of Mental Health.

* The author wrote this paper while serving as consultant to the Basic Research Program, National Institute of Mental Health, National Institutes of Health, U.S. Public Health Service, Department of Health, Education, and Welfare.

NOTATIONS

f is the fraction of the somatic cells which "survive," in the sense of remaining functional up to a given age.

f^* is the "surviving" fraction of the somatic cells at the age of death.

r is the number of inherited faults.

ρ is the number of inherited faults per chromosome.

n is the average number of faults per person in the population.

τ is the basic time interval of the aging process, defined as the average time interval between two successive aging hits suffered by the chromosomes of the somatic cell.

m is the number of pairs of homologous chromosomes of the female of the species.

x_r is the number of aging hits suffered, on the average, by the chromosomes of the somatic cells up to the age of death, by an individual who has inherited r faults.

t_r is the age, at death, of an individual who has inherited r faults.

t_0 is the life-span of the species, defined as the age at death of the genetically perfect female who did not inherit any faults.

d(theor.) is the number of deaths per year that will occur in a cohort, as given by the theory.

$d(\text{theor.})_{\max.}$ is the highest number of deaths per year that will occur in a cohort, as given by the theory.

R(theor.) is the ratio of the number of deaths per year and the maximal number of deaths per year in a cohort, as given by the theory.

R(obs.) is the ratio of the number of deaths per year and the maximal number of deaths per year in a cohort, as given by the U.S. Life Tables for white females.

σ is the standard deviation of the Gaussian that approximates the observed distribution of the number of deaths per year, between the ages of 70.5 and 90.5.

τ/t_0 is the specific life-shortening per fault of the species.

δ is the life-shortening per rep for a population that has been exposed to ionizing radiation.

δ^* is the life-shortening per rep of the adults in the offspring of a population that has been exposed to ionizing radiation.

μ_1 is the spontaneous mutation rate of the haploid set of vegetative genes per generation.

μ_t is the total spontaneous mutation rate of all genes in the haploid set.

N_1 is the haploid number of vegetative genes of the species.

N_t is the haploid number of all genes of the species.

g is the postulated number of "segments" per chromosome.

NOTES ADDED IN PROOF

1. When we say that an aging hit "destroys" a chromosome of the somatic cell, we mean that that chromosome has been rendered inactive as far as its vegetative functions are concerned, i.e., the genes which the chromosome contains will fail to produce the corresponding gene products. The question whether the chromosome is inactivated in any other sense is left open for the present. Thus it is left open whether, if a cell containing an inactivated chromosome were to duplicate, the inactivated chromosome would or would not duplicate and whether or not it would remain inactive after such a duplication. One might, for instance, imagine that the chromosomes of the somatic cell contain NNA strands which fulfil a vegetative function in the somatic cell by producing the specific gene products but do not duplicate when the cell duplicates. Aging hits would then inactivate these vegetative "copies" rather than the genetic copies. The latter would duplicate when the cell duplicates and would then produce fresh vegetative copies. This is just one of several assumptions which one may make concerning the nature of the aging hits. For the present, we are free to choose among several such *ad hoc* assumptions.

2. In the case of exposed animals it is conceivable that their life is shortened, not only through the induction of gene mutations in the chromosomes of their somatic cells by the ionizing radiation, but perhaps also through some other effects of the ionizing radiation on their somatic cells, which may involve the chromosomes or some other components of the cell. Among such effects might be the breakage of chromosomes which may lead to the loss of a chromosome. However, the theory here presented does not cover the life-shortening effect of ionizing radiation which is due to causes other than the induction of gene mutations in the somatic cells of the chromosomes. Disregarding such other effects, the "surviving" fraction of the somatic cells of an exposed female may be computed on the basis of the faults induced in the chromosomes of her somatic cells by the ionizing radiation. For a genetically perfect female who is exposed to a dose of ionizing radiation which induces, on the average, p faults per somatic cell, we may write for the "surviving" fraction of somatic cells:

$$f = [2e^{-\xi}e^{-p/2m}(1 - e^{-\xi}) + e^{-2\xi}]^m,$$

for $p/2m \ll 1$ and $p/x \ll 1$ we may write, in analogy to (13)

$$\ln \frac{1}{f^*} = \frac{1}{4m}\left(\frac{t_p}{\tau} + p\right)^2\left[1 - \frac{1}{2m}\left(\frac{t_p}{\tau} + p\right)\right],$$

where t_p is the age of death of a genetically perfect female exposed to a dose of ionizing radiation that induced an average of p faults in the chromosomes of her somatic cells.

ON IDENTICAL VANISHING OF HOLOMORPHIC FUNCTIONS IN SEVERAL COMPLEX VARIABLES*

By S. Bochner

MATHEMATICS DEPARTMENT, PRINCETON UNIVERSITY

Communicated November 7, 1958

We will make some comment on the following recent result of J. J. Kohn.[1] In the complex space $C_n:(z_1, \ldots, z_n)$ if $f_1, \ldots, f_n$ are holomorphic in the unit ball and have continuous boundary values on the boundary, and if on the boundary the linear combination $\bar{z}_1 f_1 + \ldots + \bar{z}_n f_n$ is 0, then the functions $f_1, \ldots, f_n$ are identically 0.

1. We denote by D any bounded *circular* domain in C_n, by B its boundary, and by B_0 an open neighborhood in B. We do not assume that D contains the origin in C_n, but only that B_0 has a positive distance from it. For any integer m, $m \leqq n$, we take functions $f_1, \ldots, f_m$ which are holomorphic in D and have continuous boundary values on B_0. We also take m polynomials $P_1, \ldots, P_m$ in n symbols $\zeta_1, \ldots, \zeta_n$, homogeneous and of a common degree and we denote the degree by r. Finally, we introduce the function

$$\Phi(\zeta; z) = \sum_{\mu=1}^{m} P_\mu(\zeta) f_\mu(z) \tag{1}$$

in the $2n$ variables $\zeta_1, \ldots, \zeta_n; z_1, \ldots, z_n$. If for $z \in D$ we make the specialization $\zeta_\nu = \bar{z}_\nu$, $\nu = 1, \ldots, n$, then we obtain a function $\Phi(\bar{z}; z)$ which in D is analytic in the $2n$ real variables $x_1, \ldots, x_n; y_1, \ldots, y_n$, $z_\nu = x_\nu + iy_\nu$, and has continuous boundary values on B_0.

THEOREM 1. (i) *If the function $\Phi(\bar{z}; z)$ is 0 on B_0, then the function $\Phi(\zeta; z)$ is identically 0 in $(\zeta; z)$.*

(ii) *Hence, if the polynomials $P_1, \ldots, P_m$ are linearly independent over constants, then $f_1, \ldots, f_m$ are all identically 0.*

Proof: In the case considered by Kohn, $P_1 = \zeta_1, \ldots, P_n = \zeta_n$. These polynomials are independent and his conclusion follows.

In our general case, let $a = (a_1, \ldots, a_n)$ be a fixed point of D, not the origin in C_n. If we put $z_\nu = ta_\nu$, $\nu = 1, \ldots, n$, then we have

$$\Phi(\bar{t}\bar{a}; ta) = \bar{t}^r \sum_{\mu=1}^{m} P_\mu(\bar{a}) f_\mu(ta), \tag{2}$$

and in the complex t-plane the latter sum is defined and holomorphic in a certain annulus with center at $t = 0$. Because of our assumptions, there is a neighborhood D_0 in D such that for each fixed $a \in D$, the sum in equation (2) is 0 on an arc of the t-annulus. Hence the sum is identically 0 in t, and hence for $z \in D_0$, $\Phi(\bar{z}; z)$ is 0. By a uniqueness theorem of E. Cartan, the function $\Phi(\zeta; z)$ is identically 0 as claimed.

2. In the case of the unit ball a statement can be made even if the polynomials $P_1, \ldots, P_n$ are arbitrary non-homogeneous, and the reader will readily verify that this statement subsumes Theorem 1, for such a D.

THEOREM 2. *If D is the unit ball, and $P_1, \ldots, P_m$ are any polynomials in $\zeta_1, \ldots, \zeta_n$, then the vanishing of $\Phi(z; z)$ on a neighborhood B_0 of the boundary of D implies a representation*

$$\Phi(\zeta; z) = (1 - \zeta_1 z_1 - \ldots - \zeta_n z_n)\psi(\zeta; z) \tag{3}$$

in which $\psi(\zeta; z)$ is a polynomial in (ζ_ν) with coefficients which are holomorphic in D.

Proof. If we put

$$u = 1 - \bar{z}_1 z - \ldots - \bar{z}_n z, \tag{4}$$

then D is defined by $u > 0$ and B by $u = 0$. If now we substitute

$$\bar{z}_1 = \frac{1}{z}(1 - u - \bar{z}_2 z_2 - \ldots - \bar{z}_n z_n)$$

in $\Phi(\bar{z}; z)$ we obtain an expression $A(u, \bar{z}_2, \ldots, \bar{z}_n; z_1, \ldots, z_n)$ which is a polynomial in $u, \bar{z}_2, \ldots, \bar{z}_n$ with coefficients in $z_1, \ldots, z_n$ which are holomorphic in a neighborhood D_0 in D bordering on an open part of B_0. For $u = 0$ the expression has value 0. Hence we can put

$$A(u, \bar{z}_2, \ldots, \bar{z}_n; z) = uB(u, \bar{z}_2, \ldots, \bar{z}_n; z).$$

and if herein we replace u by the sum (4), we obtain a representation (3) to the following extent. The object (3) is a polynomial in (ζ_ν) with coefficients which are holomorphic in D_0, and the equality (3) is valid for $z \in D_0$ and $\zeta_\nu = \bar{z}_\nu$. However, we can form the quotient

$$\psi(\bar{z}; z) = \frac{\Phi(\bar{z}; z)}{1 - \bar{z}_1 z_1 - \ldots - \bar{z}_n z_n}$$

in the entire unit ball D, and it is analytic in the real variables $(x_\nu; y_\nu)$. Therefore, $\psi(\zeta; z)$ has a holomorphic continuation from D_0 into D, which completes the proof of the theorem.

* This research was supported by the United States Air Force through the Office of Scientific Research of the Air Force Research and Development Command.

[1] J. J. Kohn, "A Boundary Condition for the Vanishing of n Holomorphic Functions in Complex n-Space," *Proc. Am. Math. Soc.*, **9**, 175–177, 1958.

---

*LINEAR AND ALGEBRAIC DEPENDENCE OF FUNCTIONS ON COMPACT COMPLEX SPACES WITH SINGULARITIES**

BY S. BOCHNER

MATHEMATICS DEPARTMENT, PRINCETON UNIVERSITY

Communicated November 20, 1958

We will axiomatize the following situation. Take a (non-compact) complex space W^m, entirely regular, and form the quotient space $V = W^m/R$ relative to some equivalence relation R, say relative to the action of a group of homeomorphisms. It is the space V we are interested in. We assume that it is compact, but about its "regularity" (or what may be left of it) we only assume as follows. For some complex dimension n, $n \leq m$, there is in W^m a finite number of separate complex n-cells, each holomorphically immersed, such that the union of their pro-

jections covers V. In the axiomatization we eliminate V altogether, and only retain the separate n-cells and a "weak" structure to bind them, as suggested by R.

In the space of n complex variables $z = (z^1, \ldots, z^n)$ we take a finite number of neighborhoods U_a, $a = 1, \ldots, A$, each represented by the unit ball $\|z_a\| < 1$, and, for every a, some subneighborhood T_a, as represented by $\|z_a\| < r_a$, $r_a < 1$. With each z_a in each U_a, we associate permanently some point z_b of some T_b ($b \neq a$ or $b = a$), in symbols

$$z_b = \phi(z_a). \tag{1}$$

That is all. This is our (compact) pseudo-space, and we denote it by S.

1. A *scalar* function on S is a set $f = (f_1, \ldots, f_A)$ such that $f_a(z_a)$ is an ordinary point function on V_a and that equation (1) implies

$$f_a(z_a) = f_b(\phi(z_a)). \tag{2}$$

Our first statement is as follows. *If each $f_a(z_a)$ is holomorphic, then each is a constant C_a* (maximum modulus theorem).

2. More generally, an *automorphic* (or density) function arises if for every $z_a \in U_a$, $a = 1, \ldots, A$, there is given a *fixed* number $\lambda(z_a)$, with a joint bound, $|\lambda(z_a)| < M$, and if we replace equation (2) by

$$f_a(z_a) = \lambda(z_a) \cdot f_b(\phi(z_a)). \tag{3}$$

The automorphic functions $f = (f_1, \ldots, f_A)$ constitute, in an obvious manner, a vector field over complex constants and the next statement is as follows:[1] *For holomorphic automorphic functions this vector space has a finite basis* (linear dependence of automorphic functions).

3. Even more generally we may assume that the "automorphy factor" $\lambda(z_a)$ is not a number but a N by N matrix of numbers, and that each $f_a(z_a)$ has N components, so that equation (3) is an affine transformation between such components. The object $(f_1, \ldots, f_A)$ is then representative of a general tensor (linear vector bundle over S), and the following statement still remains:[2] *Any vector field of holomorphic tensors of a given type has a finite basis.*

4. Finally, we consider meromorphic functions $f = (f_a)$. Each $f_a(z_a)$ is a quotient $p_a(z_a) : q_a(z_a)$ and there is a system of factors $\lambda(z_a)$ which depends on the given function, again with $|\lambda(z_a)| < M$, such that

$$p_a(z_a) = \lambda(z_a) p_b(\phi(z_a)), \quad q_a(z_a) = \lambda(z_a) q_b(\phi(z_a)).$$

Then the following theorem holds:[3] *To any $n + 1$ meromorphic functions $f^1, \ldots, f^{n+1}$ there is a polynomial $P(\zeta^1, \ldots, \zeta^{n+1}) \not\equiv 0$ such that $P(f^1, \ldots, f^{n+1}) \equiv 0$* (algebraic dependence of meromorphic functions).

There are versions of this theorem on other types of complex spaces with singularities.[4]

5. Finally, we note that in all preceding statements the ordinary regular neighborhoods U_a can be replaced by more general local varieties which are themselves already singular. All that is needed is that, for holomorphic functions suitably defined on them, a suitable version of the classical Schwarz's Lemma shall continue to hold. Such generalizations are indeed available.[5]

* This work was supported by the United States Air Force through the Office of Scientific Research of the Air Research and Development Command.

[1] S. Bochner, "Algebraic and Linear Dependence of Automorphic Functions in Several Variables," *J. Indian Math. Soc.*, **16**, 1–6, 1952.

[2] S. Bochner, "Tensor Fields with Finite Basis," *Ann. Math.*, **53**, 402–411, 1951.

[3] J. P. Serre, "Fonctions automorphes," *Séminaire École Norm. Sup.*, **2**, (Paris 1953/4); C. L. Siegel, "Meromorphe Funktionen auf kompakten analytischen Mannigfaltigkeiten," *Nachr. Akad. Wiss. Göttingen*, **2a**, 71–77, 1955.

[4] R. Remmert, "Meromorphe Funktionen in kompakten komplexen Räumen," *Math. Ann.*, **132**, 277–288, 1956.

[5] S. Bochner, and W. T. Martin, "Complex Spaces with Singularities," *Ann. Math.*, **57**, 490–519 1953.

---

*LINE-ELEMENT FIELDS ON THE TORUS**

BY BRUCE L. REINHART

UNIVERSITY OF MICHIGAN

Communicated by Solomon Lefschetz, November 10, 1958

1. A line-element field on the two-dimensional torus T^2 is a cross-section in the unit tangent bundle (or the projective space bundle if we consider unoriented line elements). Since either of these bundles is the product bundle $T^2 \times S^1$, we may consider such a field as a mapping $F: T^2 \to S^1$, which we shall assume to be of class C^1. We shall parametrize S^1 as follows: Consider T^2 as the quotient of the plane E^2 by the points with integer coordinates. At each point, let $\theta(v)$ be the angle between the vector v and the positive x axis. Then $\theta/2\pi$ will be taken as parameter on the fiber of the tangent sphere bundle and θ/π on the fiber of the tangent projective space bundle.

Let $F_*: H_1(T^2) \to H_1(S^1)$ be the induced homology mapping. Then for any closed curve C on T^2 lying in the homology class C^*, $F_*(C^*) = \mu\xi$, where ξ is the generator of $H_1(S^1)$ and μ is the winding number of C with respect to the field F. The C^1 homotopy classes of mappings F are in one-one correspondence with the induced maps F_*.[1] In what follows, we shall be interested in the relation between the integral curves of F, considered as defining a differential equation on T^2, and the map F_*.

LEMMA 1. *If C is a closed integral curve, the homology class C^* lies in the kernel of F_* and, in fact, generates it if F_* is not identically zero.*

2. All the closed integral curves of F belong to the same homology class, say C^*. If Γ^* is any other homology class containing a simple closed curve, and $\Gamma \in \Gamma^*$ is a simple closed curve, it follows from elementary properties of the intersection ring of the torus and the Poincaré-Bendixson theorem[2] that all integral curves of F meet Γ.

LEMMA 2. *There exists a C^1 simple closed curve $\Gamma \in \Gamma^*$ which has everywhere a non-zero tangent vector and is tangent to F at only a finite number of points.*

Henceforth, let us denote by Γ a curve in Γ^* having the fewest possible points of tangency. Let $\mu(\Gamma)$ be the winding number of Γ with respect to F.

THEOREM. $\mu(\Gamma) = 0$ *if and only if* $F_* \equiv 0$. *In particular, if there is no closed integral curve, these conditions hold.*

Proof: If there exists a closed integral curve, then both Γ^* and C^* are in the kernel of F_*, so $F_* \equiv 0$. If there is no closed integral curve, let Γ' be a cycle without contact.[3] Any integral curve of F meets Γ' infinitely often; let P be a point of accumulation of these intersections. We may then construct a C^1 closed curve C' which differs from the integral curve only in an arbitrarily small neighborhood of P. C' and Γ' always cross in the same direction; hence their intersection number is non-zero, so that they neither bound nor are homologous. The winding number of Γ' is zero because it is without contact, while that of C' is zero because it differs from an integral curve by a small amount. Hence $F_* \equiv 0$ as before.

COROLLARY 1. *If there exists on T^2 any curve whose winding number with respect to F is non-zero, then there exists a closed integral curve of F.*

3. Making use of the cycle Γ, we may generalize the rotation number $\lambda(\Gamma)$, which is classically defined for the case that there exists a cycle without contact.[2, 3] Let Γ and Γ_1 form a basis for $H_1(T^2)$, and let $F_*(\Gamma) = i\xi$, $F_*(\Gamma_1) = j\xi$.

PROPOSITION. *$\mu(\Gamma) = i$ and $\lambda(\Gamma) = -j/i$, the latter holding only in case $\mu(\Gamma) \neq 0$.*

By considering the number of points of tangency on the cycle Γ, we may classify qualitatively the various kinds of integral curve families of non-oriented line-element fields.[4] From this classification, the following corollaries are immediate:

COROLLARY 2. *The non-oriented line element field F is orientable if and only if i and j are both even.*

COROLLARY 3. *The number of closed integral curves is at least the greatest common divisor of i and j.*

* This research was supported by the Office of Naval Research.

[1] N. E. Steenrod, *The Topology of Fibre Bundles* (Princeton: Princeton University Press, 1951), §34.

[2] E. A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations* (New York: McGraw-Hill Book Co., 1955), chaps. xvi–xvii.

[3] C. L. Siegel, "Note on Differential Equations on the Torus," *Ann. Math.*, **46**, 423–428, 1945.

[4] H. Kneser, "Reguläre Kurvenscharen auf den Ringflächen," *Math. Ann.*, **91**, 135–154, 1924.

ON A PROBLEM OF MAZURKIEWICZ CONCERNING THE BOUNDARY OF A COVERING SURFACE

BY PETER SEIBERT

A MEMBER OF RIAS, 7212 BELLONA AVE., BALTIMORE, MD.

Communicated by Solomon Lefschetz, November 13, 1958

1. In his note[1] Mazurkiewicz showed that, by introducing a certain metric on a Riemann covering surface and completing the obtained metric space, a concept of "boundary elements" can be defined which corresponds precisely to the usual notion of transcendental singularities of meromorphic functions defined on the surface. This boundary, considered by itself, is a complete separable metric space. Moreover, Mazurkiewicz proved in the above-mentioned note that the dimension (in the sense of Menger-Urysohn) of the boundary never exceeds the value 2 and raised the question whether covering surfaces with two-dimensional boundaries exist. To our knowledge an answer to this question has, so far, not been given.

In the present note we give a simple example of a covering surface with a two-dimensional (and, moreover, compact) boundary. Furthermore, we show that the boundary of a simply connected covering surface is, at most, one-dimensional if it is locally compact.

2. Let F be a two-dimensional manifold, S the Riemann sphere and s an interior mapping[2] of F into S. In this sense we call F a *covering surface* of S and s the *projection mapping* of F.

Now consider two points $p, q, \epsilon F$ and denote by $\mathbf{C}$ the set of all Jordan arcs C on F connecting p and q. Then, if the diameter of a set in S with respect to the spherical metric is denoted by *diam*, the distance between p and q will be defined by

$$\rho(p, q) = \inf_{C \epsilon \mathbf{C}} diam\ s(C). \tag{1}$$

It has been shown in a previous paper[3] that this definition yields a metric which is compatible with the original topology of F, so that a sequence $\{p_n\}$ tends to a point p if and only if $\rho(p_n, p) \to 0$.[4] The completion of the surface F will be called the *extended surface* and denoted by F^*. Then the *boundary* $\mathbf{Z}$ of F is defined by

$$\mathbf{Z} = F^* - F.$$

Therefore, the boundary elements (which we will usually denote by ζ) correspond to the equivalence classes of those fundamental sequences of F which possess no accumulation point on F.

We now extend the projection s continuously onto the boundary, thus obtaining a mapping of the extended surface F^* into S.

3. In the present section we shall assume F to be *simply connected*. In our paper (RS) it was shown that in this case a cyclic order can be defined on the boundary. Subsequently, we established a relation between the two topologies on the boundary defined by the Mazurkiewicz metric and by the cyclic order, respectively. As a corollary of this relation the following result was obtained:[5]

LEMMA. *The boundary of a simply connected covering surface always permits a one-one continuous and order-preserving map into the circular line. (Continuity here is meant with respect to the metric on* $\mathbf{Z}$.)

THEOREM 1. *If the boundary* $\mathbf{Z}$ *of a simply connected covering surface* F *is locally compact, the condition*

$$dim\ \mathbf{Z} \leqq 1 \tag{2}$$

holds.

Proof: $\mathbf{Z}$ is a separable metric space.[6] If, in addition, it is locally compact, it can be represented in the form

$$\mathbf{Z} = \bigcup_{\nu=1}^{\infty} \Phi_\nu, \tag{3}$$

where each of the sets Φ_ν is compact.[7] The mapping mentioned in the lemma maps each Φ_ν *homeomorphically* onto a subset of the circle, so that

$$\dim \Phi_\nu \leqq 1 \quad (\nu = 1, 2, \ldots) \tag{4}$$

follows. Finally, formulas (3) and (4), together with a theorem of Hurewicz,[8] yield the required inequality (2).

Remark: If the surface F is generated by an entire function f, its boundary is totally disconnected and therefore of dimension 0. This statement is true despite the fact that some entire functions assume every complex number as an asymptotic value.[9] Indeed, owing to a well-known theorem of Lindelöf, two non-equivalent asymptotic paths of f with finite asymptotic values are always separated by paths along which f tends to infinity. In our notation[10] this means: Given two boundary points $\zeta_1, \zeta_2 \in \mathbf{Z}$ with $s(\zeta_1), s(\zeta_2) \neq \infty$, each of the sets (ζ_1, ζ_2), (ζ_2, ζ_1) contains at least one boundary point over ∞. Now let $\mathbf{A}$ be a connected subset of Z, for which $s(\mathbf{A})$ does not contain the point at infinity. Then $\varphi(\mathbf{A})$ is connected, because φ is continuous. Since φ is also order-preserving, any two points of $\varphi(\mathbf{A})$ are separated by images of points over ∞, so that $\varphi(\mathbf{A})$, and therefore $\mathbf{A}$, reduces to a single point. The set of boundary points over ∞, so that $\varphi(\mathbf{A})$, and therefore $\mathbf{A}$, reduces to a single point. The set of boundary points over ∞, on the other hand, is known to be a discontinuum.[11] Therefore, we have the following proposition: *A simply connected Riemann covering surface which does not cover a point $a \in S$ is either a hyperbolic type or its boundary is a discontinuum.*

4. THEOREM 2. *There exists a covering surface F of the sphere S whose projection mapping s maps the boundary $\mathbf{Z}$ of F homeomorphically (and, moreover, isometrically) onto S. More generally, given a closed subset A of S, there exists a covering surface whose boundary is mapped by s isometrically onto A.*

Proof: We first consider the case $A = S$. Denote by $d(a, b)$ the spherical distance between two points $a, b \in S$. Let $\{L_n\}_0^\infty$ be a sequence of sets each of which consists of a finite number of segments $L_{n,\mu}$ ($\mu = 1, 2, \ldots, m_\mu$) with endpoints $a'_{n,\mu}, a''_{n,\mu}$ and satisfies the following conditions:

(i) $L_{n',\mu'} \cap L_{n'',\mu''} = \phi$ for all n', n'', μ', μ'' with $(n', \mu') \neq (n'', \mu'')$.

(ii) $d(a'_{n,\mu}, a''_{n,\mu}) \leqq 2^{-n}\pi$ for all n, μ.

(iii) Denoting by $V_\epsilon(B)$ the ϵ-neighborhood $\{a \in S: \inf_{b \in B} d(a, b) < \epsilon\}$ of a set $B \subset S$, the relation

$$V_{2^{-n}\pi}(L_n) = S$$

holds for all n.

Example: Denote by (θ, φ) the point on the sphere of longitude θ and latitude φ ($0 \leqq \theta < 2\pi$, $-\frac{\pi}{2} \leqq \varphi \leqq +\frac{\pi}{2}$), and consider the segments $L_{n;\kappa,\lambda}$ of the parallel circles $\varphi =$ constant bounded by the points

$$a'_{n;\kappa,\lambda} = (2^{-n}\pi\lambda, \quad 2^{-n-1}\pi k), \quad a''_{n;\kappa,\lambda} = a'_{n;\kappa,\lambda} + (2^{-n-1}\pi, 0),$$

where

$$\kappa = \begin{cases} 0 \text{ for } n = 0 \\ \pm 1, \pm 2, \ldots, \pm (2^n - 1) \text{ for } n > 0, \end{cases} \qquad \lambda = 0, 1, 2, 3, \ldots, 2^{n+1}.$$

Then choose as set L_n the set $\{L_{n,\kappa,\lambda}\}$ with κ, λ varying as indicated above.

We then construct the surface F in the following manner: By removing from S the sets of segments[12] L_{n-1} and L_n, we obtain a sequence of "sheets" $F_n (n = 0, 1, 2, \ldots)$ which, by joining F_n with F_{n+1} along both edges of all slits of L_n, yields

a surface of the required type. (Here the projection s is, of course, the mapping which assigns to every point of F the point of S with the same coordinate.)

The proof will be carried out in two steps by showing that:

(α) Every point of S possesses an inverse image with respect to s in $\mathbf{Z}$: $s(\mathbf{Z}) = S$.

(β) Given two elements $\zeta', \zeta'' \in \mathbf{Z}$, the relation

$$\rho(\zeta', \zeta'') = d(s(\zeta'), s(\zeta''))$$

holds.

Given a point $a \in S$, there exists, according to (ii) and (iii), a sequence $\{p_n\}$ of branch points of F such that every p_n is an end-point of a segment of L_n and $s(p_n)$ tends to a for $n \to \infty$. We have to show that $\{p_n\}$ is convergent. The points p_{n-1} and p_n both lie on the boundary of the sheet F_n. Now denote by S_n the segment with end-points $s(p_{n-1})$, $s(p_n)$ and consider the subset

$$D_n = \{p \in F_n : s(p) \in V_{2^{1-n}\pi}(S_n)$$

of F_n. We conclude from (ii) that none of the cuts $L_{n,\mu}$ separates p_{n-1} and p_n within the set D_n, and from (i) that the same is true for any union of such segments. Therefore, the two points can be connected by an arc in D_n and, according to the definition of the distance ρ, the inequality

$$\rho(p_{n-1}, p_n) \leqq diam\, s(D_n) = d(s(p_{n-1}), s(p_n)) + 2^{2-n}\pi \tag{5}$$

holds. Since $\{s(p_n)\}$ converges, so does $\{p_n\}$, and assertion (α) follows.

To prove (β), consider two points $\zeta', \zeta'' \in \mathbf{Z}$ and two sequences of branch points $\{p_n'\}$ and $\{p_n''\}$ with ζ' and ζ'' as their respective limits[13] and assume $s(p_n')$, $s(p_n'')$ to be end-points of segments $L_{n,\mu}$. Then for every n the following pair of inequalities holds:

$$d(s(p_n'), s(p_n'')) \leqq \rho(p_n', p_n'') \leqq d(s(p_n'), s(p_n'')) + 2^{1-n}\pi.$$

The first is obvious. The second is obtained in the same way as (5). Since

$$\lim_{n \to \infty} d(s(p_n'), s(p_n'')) = d(s(\zeta'), s(\zeta'')),$$

the assertion follows.

In the case $A \subset S$ we denote by A_n the set $V_{2^{-n}\pi}(A)$ and omit in L_n all segments which are disjoint to A_n.[14] It is then easy to see that $s(\mathbf{Z})$ is precisely the set A.

5. Finally, we mention two open questions concerning the dimension of the boundary, one of which is topological, the other function theoretic:

I. Does relation (2) hold for simply connected covering surfaces in general?

II. Do there exist (concrete) Riemann surfaces with null-boundary (in the sense of Nevanlinna) whose boundary (in the sense of Mazurkiewicz) is of dimension 2?

[1] S. Mazurkiewicz, "Sur les points singuliers d'une fonction analytique," *Fund. Math.*, **17**, 26–29, 1931.

[2] I.e., a continuous mapping which transforms open sets into open sets and under which no non-degenerate connected set has a single point as its image.

[3] Cf. Seibert, "Über die Randstukturen von Überlagerungsflächen," to appear in *Mathematische Nachrichten*, **19**. This paper will be referred to as "RS" hereafter.

[4] For the pertinent arguments in the paper RS, it apparently makes no difference whether the

metric introduced there or the one defined by (1) is considered. The distance (1) is called the *distance with respect to the mapping (s, F)* by L. Cesari, who introduces it in compact metric spaces (cf. "Surface Area," *Ann. Math. Studies*, **35**, 159 (1956).

[5] Cf. RS, §7(f).

[6] Cf. *ibid.*, §3(c).

[7] Cf. C. Kuratowski, *Topologie* (Warzawa-Wrocław, 1950), II, 51, théor. 4.

[8] Cf. *ibid.*, I, 176, théor. 1; cf. also Hurewicz-Wallman, *Dimension Theory* (Princeton: Princeton University-Press, 1948), p. 30, theor. III 2.

[9] Cf. W. Gross, "Eine ganze Funktion, für die jede komplexe Zahl Konvergenzwert ist," *Math. Ann.*, **79**, 201–208, 1918.

[10] Cf. RS, §4 for the relationship between asymptotic paths and boundary points and §5 for the notation used here.

[11] Cf. ibid. §3(d).

[12] $L_{-1} = \phi$.

[13] Apparently every boundary point is the limit of a sequence of branch points. Indeed, if $\zeta \epsilon z$, either every neighborhood of ζ contains parts of infinitely many sheets, or all sufficiently small neighborhoods are contained in a single sheet F_n. In the first case ζ is a cluster point of cuts $L_{n,\mu}$ and therefore, according to (ii), also of branch points. The second case cannot occur because every boundary point of a sheet F_n is an interior point of F.

[14] We can, of course, assume A to be non-empty.

---

*THE EFFECTS OF STRONTIUM-90 ON MICE**

BY BARCLAY KAMB AND LINUS PAULING

CALIFORNIA INSTITUTE OF TECHNOLOGY, PASADENA, CALIFORNIA

Communicated November 17, 1958

On Sept. 19, 1958 there was published in *Science* a paper by Dr. Miriam P. Finkel of Argonne National Laboratory in which she communicated her observations on the effects of strontium-90 injected into mice on life expectancy and on incidence of tumors of bone and blood-forming tissues.[1] She discussed the question of whether or not the effects are proportional to the amount of injected strontium-90 at low doses, and reached the conclusion that it is likely that there is a threshold with value for man between 5 and 15 μc. (as compared with the present average value from fallout, about 0.0002 μc., and the predicted steady-state value from fallout for testing of nuclear weapons at the average rate for the past five years, about 0.02 μc.). Her paper ends with the sentence "In any case, the present contamination with strontium-90 from fallout is so very much lower than any of these levels that it is extremely unlikely to induce even one bone tumor or one case of leukemia."

On the same day, Sept. 19, 1958, newspapers throughout the United States published accounts of this work. For example, the Pasadena (Calif.) *Star-News* contained an article with the headline "Tests on Mice Show Fallout Safe" and the first sentence, "A woman researcher says tests on mice show that the present fallout from nuclear weapons tests will not produce a single case of bone cancer or leukemia in humans." The *New York Times* published accounts of the work on both September 19 and September 28.

We have made an analysis of Dr. Finkel's data that shows that she had no justi-

fication whatever for her concluding statement. All of her data are compatible with a zero threshold for strontium-90. Moreover, the statistical analysis shows that in order for Dr. Finkel to have been justified with 90 per cent confidence (10 per cent type-II error) in making her concluding statement on the basis of her data she would have to have used over 1,000,000,000,000 mice in each of her groups, instead of the 150 or less that were used. It is hard for us to understand how such a serious error could be made in Dr. Finkel's argument, leading her to publish her seriously misleading statement about this matter of great importance.

The Mice Experiments.—In the studies described by Dr. Finkel young adult female mice (strain CF No. 1, about 70 days old) were given a single intravenous injection of an isotonic equilibrium mixture of strontium-90 chloride and yttrium-90 chloride. There were twelve injected groups, ranging in size from 150 mice for the group receiving the smallest amount (1.3 μc/kg body weight) to 15 for that receiving the largest amount (9330 μc/kg, an amount that caused death of about 50 per cent within 30 days). The control group contained 150 mice. The author states that there is 11 per cent retention (at 600 days) of the injected radioactive material. Report was made of the fractional decrease in average survival time, the incidence of animals with osteogenic sarcomas (among 150-day survivors), and the fractional decrease in time to a 20 per cent incidence of reticular tissue tumors compared with the 20 per cent incidence time of the controls.

Studies of this sort may be of great value in providing information about the probable amount of damage done to human beings by exposure to high-energy radiation, such as that from strontium-90 produced by nuclear weapons. It is important that the analysis of the experimental results be carried out correctly. We have found that in the treatment of problems of this sort the assumption that the probability of damage is strictly proportional to the amount of radiation exposure does not in general require that a response such as decrease in life expectancy be linear, except over a very small range. Moreover, we have found that this assumption together with the theory of statistics can be applied in a reasonably straightforward way in the discussion of data such as those obtained by Dr. Finkel, as shown in the following sections.

Analysis of the Experimental Data on Life Shortening.—Our analysis proceeds from the hypothesis, induced by Lewis[2] as a result of his study of the incidence of leukemia, that exposure of the bone marrow of an animal to radiation results in an increase in the probability per unit time that the animal will die at any time thereafter, the increase being proportional to the quantity of radiation absorbed. We shall suppose that this hypothesis applies to all of the radiation-induced effects in Dr. Finkel's experiments with mice.

Let N_0 be the number of animals at the beginning of a given experiment, taken to be at $t = 0$, and let $N(t)$ be the expected number (average for many experiments of the same kind) at the later time t. Further, let $N^0(t)$ be the expected number in a "control" experiment in which no strontium-90 is injected, so that

$$g(t) \equiv -\frac{1}{N_0}\frac{dN^0}{dt}$$

is the natural specific death-rate function. We denote by α the quantity of strontium-90, in μc/kg body weight, that is retained in the animals. Then our hypothesis yields the equation

$$\frac{dN}{dt} = -N\beta\alpha t - \frac{N}{N^0(t)} \cdot N_0 g(t) \tag{1}$$

where β is a constant of proportionality relating the quantity of strontium-90 retained to the increased probability per unit time that the animals will die, this probability of course increasing linearly with time owing to the nearly constant irradiation by the decaying strontium-90.

Let $n^0(t) = N^0(t)/N_0$, so that $g(t) = -\dot{n}^0(t)$. Then on integrating equation (1) we obtain

$$\begin{aligned} N &= N_0 \exp\left[-\frac{1}{2}\beta\alpha t^2 - \int_0^t \frac{\dot{n}^0(t)}{n^0(t)}\, dt \right] \\ &= N_0 n^0(t) e^{-(1/2)\beta\alpha t^2} = N^0(t) e^{-(1/2)\beta\alpha t^2} \end{aligned} \tag{2}$$

To compare this result with the experimental data we calculate Δ, the fractional decrease in life expectancy (fractional decrease in average survival time after injection),

$$\Delta = \frac{t_0 - t_\alpha}{t_0}$$

where t_α is the life expectancy for a retained quantity α of strontium-90,

$$t_\alpha = \frac{1}{N_0} \int_0^\infty N(t) dt \tag{3}$$

The equation for Δ is

$$\Delta = 1 - \frac{1}{t_0} \int_0^\infty n^0(t) e^{-\gamma t^2}\, dt \tag{4a}$$

$$= \frac{1}{t_0} \int_0^\infty \left(t - \frac{1}{2}\sqrt{\frac{\pi}{\gamma}}\, \mathrm{Erf}\, t\sqrt{\gamma} \right) g(t)\, dt \tag{4b}$$

where for simplicity we put $\gamma = \frac{1}{2}\alpha\beta$. The result in equation (4b) is obtained by an integration by parts, and the error function is defined as

$$\mathrm{Erf}\, x = \frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2}\, dy$$

If normally (for $\alpha = 0$) all animals lived to the age t_0 and then died, so that $g(t)$ were a delta function $\delta(t - t_0)$, then we would have simply

$$\Delta = 1 - \frac{\sqrt{\pi}}{2} \frac{\mathrm{Erf}\, t_0\sqrt{\gamma}}{t_0\sqrt{\gamma}} \tag{5}$$

However. the actual lifetimes scatter with sizable dispersion about t_0. The extent of this dispersion can be estimated from the acceptance region $\Delta \leq 0.07$ quoted by Dr. Finkel (Figs. 3 and 4) as appropriate to a test of the hypothesis of no difference between the responses of the control population and of a population injected with a given dose of strontium-90. If we assume (1) that the test of no difference applies to Curve A, the life-shortening data, (2) that the test was one-tailed, and (3) that the test accounted for the uncertainty in the mean lifetime of the control population and for the uncertainty in the mean lifetime of her group 10, the highest-dosed population to fall within the acceptance region (except for the "peculiar result" for group 8), then we find that the estimated standard deviation for $g(t)$ is $\hat{\sigma} = 258$ days.

These assumptions are somewhat uncertain, as explained later, but they are the best that can be made from the information given in Dr. Finkel's paper. The uncertainty in drawing any conclusions about $g(t)$ from Dr. Finkel's data lead us to take a more general approach. Gompertz discovered that for animal populations the logarithm of the "age-specific death rate" is closely a linearly increasing function of time. For man the age-specific death-rate doubling time is about 8 years. Jones[3] has pointed out that the doubling times for different animal species are approximately proportional to the mean life spans for the species. We shall use this information to derive a hypothetical death-rate function $g(t)$ for the mouse population used in Dr. Finkel's experiments.

The Gompertz law is

$$\ln g(t) = C + Bt \tag{6}$$

which yields

$$n^0(t) = \exp\left[-A(e^{Bt} - 1)\right] \tag{7}$$

where A $(= e^C/B)$ is a constant and where B is related to the doubling time τ_D by

$$B = \frac{\ln 2}{\tau_D} \tag{8}$$

A is to be chosen so as to give the correct mean life span:

$$t_0 = \int_0^\infty \exp\left[-A(e^{Bt} - 1)\right]dt = \frac{e^A}{B}\left[-\mathrm{Ei}(-A)\right] \tag{9}$$

The exponential integral $\mathrm{Ei}(x)$ is defined by Jahnke and Emde.[4] If $t = 0$ is taken to be a time shortly after birth, but long enough after birth to exclude infant deaths (which are omitted in Gompertz' treatment), then t_0 is T, the mean life span from birth to death. If then T/τ_D is a constant for all animal species, we find from equations (8) and (9) that A is a constant, independent of species, given by the solution of the equation

$$e^A\left[-\mathrm{Ei}(-A)\right] = \frac{T}{\tau_D}\ln 2 \tag{10}$$

Assuming $T = 60$ years for man, with $\tau_D = 8$ years, we find $A = 0.0032$. The solution of equation (9) is obtained with the help of the expansion[4]

$$-\mathrm{Ei}(-x) = -\ln\Gamma x + x - \frac{x^2}{2!2} + \ldots \tag{11}$$

where $\Gamma = 1.781$.

The death-rate function is

$$g(t) = AB \exp [Bt - A(e^{Bt} - 1)]$$

The dispersion of life spans is measured by the standard deviation σ of $g(t)$:

$$\sigma^2 = \int_0^\infty t^2 g(t)dt - t_0^2 = 2\int_0^\infty t n^0(t)dt - t_0^2$$

The second expression results from an integration by parts. A numerical integration is required to obtain σ^2, which can most easily be carried out with $n^0(t)$ values from equation (7). In this way we find from equations (9) and (10), with $t_0 = T$, that

$$\sigma = 0.24T$$

For Dr. Finkel's mice, reported to have $T = 670$ days, we have $\sigma = 161$ days, in rather poor agreement with the value $\sigma = 258$ days inferred above from her paper.

In the calculations that follow we have used equations (7) and (10), with the assumption $T = t_0 = 600$ days, although actually the mice were about 70 days old at the beginning of the experiment. Thus we have used a doubling time τ_D of 80 days, and our $g(t)$ has standard deviation $\sigma = 141$ days. The assumption $\tau_D = 80$ days agrees with the value quoted by Jones[3] for mice. Values of $n^0(t)$ for these parameters are given in Table 1. The difference between assuming $T = 600$ days and assuming $T = 670$ days is not great; in fact, survival curves calculated from equation (5), which assumes $\sigma = 0$, do not differ greatly from curves obtained by the more refined procedure that we have used.

TABLE 1

t (Days)	$n^0(t)$	$e^{-\mu t^2}$
0	1.000	1.000
80	0.997	0.997
160	0.990	0.987
240	0.978	0.972
320	0.953	0.950
400	0.908	0.923
480	0.818	0.891
560	0.666	...
640	0.443	...
720	0.195	...
800	0.038	...
880	0.000	...

We proceed now to compare equation (4a), evaluated with the help of equations (7) and (10), with the experimental life-shortening data. We assume with Dr. Finkel that $\alpha = 0.11\alpha^*$ where α^* is the injected dose of strontium-90 in μc/kg body weight, and we attempt to choose the available parameters so as best to reproduce the observed life-shortening data $\Delta(\alpha^*)$. There are two parameters: the constant β, and the no-dose life-shortening Δ_0, the latter arising from the fact that we cannot give great weight to Dr. Finkel's zero point because of the statistical uncertainty

in the observed mean life span of the control population. Thus the theoretical curve to be fitted to the data is

$$\Delta(\alpha^*) = \Delta_0 + 1 - \frac{1}{t_0}\int_0^\infty n^0(t)e^{-(0.11/2)\beta\alpha^* t^2}\,dt \tag{12}$$

The inclusion of Δ_0 simply as a constant in equation (12) is not strictly correct: the additive term arising from an adjustment of the zero point should be written $\Delta_0(\alpha^*)$, where $\Delta_0(\alpha^*)$ is a somewhat complicated decreasing function of α^* that tends to zero as $\alpha^* \to \infty$. For simplicity, however, we ignore this complication, which proves to be unimportant for the lower radiation levels ($\alpha^* \lesssim 1000$ μc/kg), and which in any case does not much change the results obtained, because Δ_0 is small.

For very small values of α^* equation (12) reduces to

$$\Delta(\alpha^*) = \Delta_0 + \alpha^*\left[\frac{1}{2}(0.11)\beta\int_0^\infty n^0(t)t^2dt\right]$$

$$= \Delta_0 + \alpha^*\cdot\left.\frac{d\Delta}{d\alpha^*}\right|_{\alpha^*=0} \tag{13}$$

This is the linear response region. We can therefore choose preliminary values of Δ_0 and β by estimating a linear fit to the experimental points at low values of α^*. The integration in equation (13) is performed numerically, with use of equation (7).

We have calculated theoretical curves from equation (12) in three steps: (1) For $\alpha^* \leq 50$ μc/kg equation (13) applies; (2) for selected values of α^* in the range $50 \leq \alpha^* \leq 1000$ we carry out the integration in equation (12) numerically, using time intervals $\Delta t = 80$ days; (3) for $\alpha^* > 1000$ it is found that the asymptotic form of equation (12) is valid:

$$\Delta(\alpha^*) = \Delta_0 + 1 - \frac{1}{t_0}\sqrt{\frac{\pi}{0.11\beta\alpha^*}} \tag{14}$$

Because we wish to examine the result statistically, we adjust the parameters by a weighted least squares procedure. We calculate two theoretical curves $y = f_0 + f(x, \beta_1)$ and $y = f_0 + f(x, \beta_2)$ (here y is fractional life shortening, Δ; x is injected dose, α^*; and f_0 is the constant Δ_0) for two nearly correct values β_1 and β_2 of the parameter β. We ask for values $\hat{f}_0 = f_0 + \Delta f_0$ and $\hat{\beta} = \beta_1 + (\beta_2 - \beta_1)\delta$ of the parameters such that the weighted sum of the squares of the differences between the experimental values y_i and the theoretical values $y(x_i)$ is a minimum:

$$\sum_i w_i(y_i - f(x_i, \hat{\beta}) - \hat{f}_0)^2 = \min \tag{15}$$

Since $\beta_2 - \beta_1$ is small we can assume that

$$f(x_i, \hat{\beta}) = f(x_i, \beta_1) + [f(x_i, \beta_2) - f(x_i, \beta_1)]\delta$$

$$\equiv f(x_i, \beta_1) + \Delta f_i\delta \tag{16}$$

The parameter adjustments Δf_0 and δ are then given by

$$\Delta f_0 = \frac{(\sum w_i \Delta y_i)(\sum w_i \Delta f_i^2) - (\sum w_i \Delta y_i \Delta f_i)(\sum w_i \Delta f_i)}{(\sum w_i)(\sum w_i \Delta f_i^2) - (\sum w_i \Delta f_i)^2} \tag{17}$$

$$\delta = \frac{(\sum w_i \Delta y_i \Delta f_i)(\sum w_i) - (\sum w_i \Delta y_i)(\sum w_i \Delta f_i)}{(\sum w_i)(\sum w_i \Delta f_i^2) - (\sum w_i \Delta f_i)^2}$$

where $\Delta y_i = y_i - f_0 - f(x_i, \beta_1)$.

The weights w_i appearing in equations (15) and (17) should be inversely proportional to the *a priori* variances of the experimental values y_i. We take the variances to be inversely proportional to the number of animals in each experimental group. This ignores the effect of radiation in changing the dispersion of life spans, but a detailed examination shows it to be a not unreasonable procedure.

The theoretical curve obtained in the above manner is shown in Figure 1, with

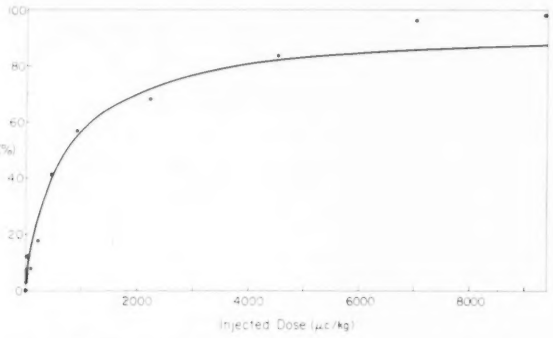

FIG. 1.—Percentage decrease in life expectancy, Δ, as a function of injected dose α^* of strontium-90. Solid curve is the theoretical curve calculated from equation (12). Solid circles are the experimental values reported by Dr. Finkel.

the experimental points for comparison. The two points at highest radiation levels lie well above the curve, doubtless because the mechanism of life shortening at the high radiation levels departs from what we have assumed, owing to the importance of subacute and acute irradiation disease, which Dr. Finkel reports to be the primary cause of death at injected doses above 2200 μc/kg. These points make little contribution to the least squares parameter adjustment, owing to their low weights, and can be omitted without sensibly changing the result. The parameters obtained are $\beta = 1.8 \times 10^{-7}$ day^{-2} (μc/kg retained)$^{-1}$, and $\Delta_0 = 2.5$ per cent.

The least-squares-fitted curve can be used to estimate the death-rate standard deviation σ: for experimental Δ values of unit weight (taken here to be for the control group and "group 12"), the estimated variance of the experimental Δ values is

$$^2 = \frac{1}{m-2} \sum_{i=1}^{m} w_i [y_i - \hat{f}_0 - f(x_i, \hat{\beta})]^2 \tag{18}$$

where m is the number of experimental points. Equation (18) takes into account the two-parameter adjustment. If M is the number of animals in groups having unit weight, then

$$\hat{\sigma}^2 = M\hat{\sigma}_\Delta{}^2 \tag{19}$$

From equations (18) and (19) we find $\hat{\sigma} = 191$ days in case the two highest points mentioned above are omitted ($\hat{\sigma} = 222$ days in case they are included). Comparing the value 191 days with the value $\sigma = 161$ days based on the Gompertz death-rate curve (670-day life span) and the value $\hat{\sigma} = 258$ days inferred from Dr. Finkel's data, we see that the theoretical curve fits the experimental values about as well as would be expected from the Gompertz curve, and somewhat better than would have been expected on the basis of Dr. Finkel's acceptance region.

In Figure 2 there is shown the portion of the theoretical curve for the lower radi-

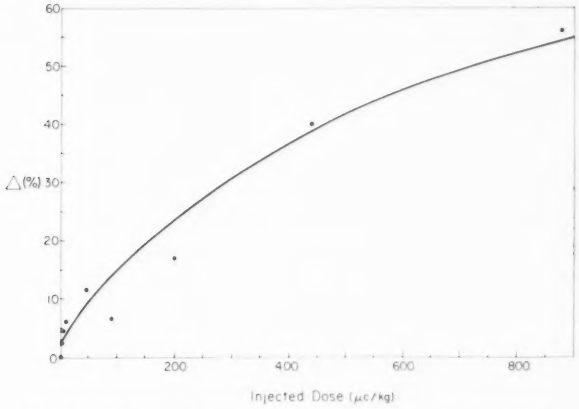

FIG. 2.—Percentage decrease in life expectancy in the low-dose region of Dr. Finkel's experiments. Theoretical curve and experimental points as in Fig. 1.

ation levels. The curvature is pronounced, and the linear response region is restricted to injected doses less than about 50 μc/kg. Most of Dr. Finkel's experiments were carried out in the nonlinear portion of the curve.

It is interesting to compare the above analysis with an alternative one based on the approach developed by Jones,[3] in which the effect of a given exposure of an animal to radiation is regarded as equivalent to an increase in physiological age of the animal by an amount proportional to the amount of radiation received. In terms of the Gompertz formulation of the natural death rate, this results in the case we are considering to the addition of a linear term in t to equation (6):

$$\ln\left(-\frac{1}{N}\frac{dN}{dt}\right) = C + Bt + B\eta\alpha t \tag{20}$$

The constant $B\eta$ in this treatment plays a role analogous to the constant β used previously.

From equations (3) and (20) we obtain

$$t_\alpha = \frac{-\mathrm{Ei}\left(-\dfrac{A}{1+\eta\alpha}\right)}{B(1+\eta\alpha)}\, e^{A/(1+\eta\alpha)} \tag{21}$$

where $A = e^C/B$ as before. In practice A is so small that the exponential integral can adequately be approximated by the logarithmic term in equation (11):

$$-\mathrm{Ei}\left(-\frac{A}{1+\eta\alpha}\right) = \ln\frac{1}{\Gamma A} + \ln(1+\eta\alpha) \tag{22}$$

Recognizing from equations (21) and (22) that

$$t_0 = \frac{1}{B}\ln\frac{1}{\Gamma A}$$

and making use of equation (8), we obtain

$$\Delta = 1 - \frac{1+\epsilon\ln(1+\eta\alpha)}{1+\eta\alpha} \tag{23}$$

where $\epsilon = \tau_D/(t_0 \ln 2) \cong 1/5$.

By choosing $\eta = 0.014$ (μc/kg retained)$^{-1}$, and by adjusting the zero point slightly as done previously, we calculate from equation (23) a theoretical curve that matches closely the curve calculated from equation (12), which is the curve shown in Figures 1 and 2. The discrepancy in Δ between the two curves is less than 0.02 over the range $0 \leq \alpha^* \leq 3000$ μc/kg, and increases to 0.04 at $\alpha^* = 9000$ μc/kg. The two curves fit the experimental data equally well, as shown by the estimates $\sigma = 191$ days for the curve from equation (12) and $\sigma = 189$ days for equation (23), calculated by the weighted sum-of-squares procedure described previously. Life-shortening data, at least of the accuracy involved here, are therefore unable to discriminate between the two analytical approaches.

Analysis of Incidence of Leukemia and Related Diseases.—The experimental data on the incidence of diseases of the blood and blood-forming tissues can be analyzed in the framework of the above treatment. However, because of the peculiar form in which the experimental results are presented ("Curve C: percentage decrease in time to a 20 per cent incidence of reticular tissue tumors compared with the 20 per cent incidence time of the controls"), the analysis is subject to greater uncertainties and difficulties and the data cannot so readily be evaluated statistically as those for the decreased life expectancy. We therefore content ourselves with a somewhat sketchy treatment, which should suffice to indicate the general nature of the problem.

Let $\lambda(t, \alpha)$ be the expected number of deaths due to these diseases that have occurred by the time t in a population having retained body burden α of strontium-90. We may then expect to find a death-rate probability parameter β_l for these diseases such that the death rate is

$$\frac{d\lambda}{dt} = N(t)\beta_l\alpha t + \frac{d\lambda_0}{dt}\frac{N(t)}{N^0(t)}$$

where $\lambda_0(t)$ is the number of deaths due to these diseases expected in the control population. To carry the analysis further we need to know the function $\lambda_0(t)$, but unfortunately Dr. Finkel presents no data that enable us to determine it. Of the various assumptions that could be made, we have chosen to assume that the nat-

ural deaths due to leukemia are distributed as though they were radiation-induced according to the same model as the deaths due to radiation from strontium-90. The natural leukemia death rate will then be equivalent to a "background" body burden α_0 of strontium-90, and equation (20) becomes

$$\frac{d\lambda}{dt} = N(t)\beta_l(\alpha_0 + \alpha)t$$

Obtaining $N(t)$ from equation (2), we have

$$\frac{\lambda(t)}{N_0} = \beta_l(\alpha_0 + \alpha)\int_0^t n^0(t)e^{-(1/2)\beta\alpha t^2}\, t\, dt$$

The expected 20 per cent incidence time τ is then the implicit solution of

$$0.11\beta_l(\alpha_0^* + \alpha^*)\int_0^\tau n^0(t)e^{-(0.11/2)\beta\alpha^* t^2}\, t\, dt = \frac{1}{5} \tag{24}$$

and the expected no-dose 20 per cent incidence time τ_0 is given by

$$0.11\beta_l\alpha_0^*\int_0^{\tau_0} n^0(t)t\, dt = \frac{1}{5} \tag{25}$$

τ_0 as given by equation (25) is not necessarily the same as the 20 per cent incidence time $\tau_0' = 565$ days observed for the control population.

To compare the theory with the experimental data we calculate from equation (24) the fractional decrease function $1 - \tau(\alpha^*)/\tau_0'$. An adequate approximate calculation for values of τ less than about 450 days ($1 - \tau/\tau_0' > 0.20$) can be made by approximating $n^0(t)$ by a Gaussian $e^{-\mu t^2}$, as shown in Table 1. In this case equation (24) becomes

$$0.11\beta_l(\alpha_0^* + \alpha^*) = \frac{\mu + \frac{1}{2}(0.11)\beta\alpha^*}{5\left[1 - \exp\left(-\tau^2\left(\mu + \frac{1}{2}0.11\beta\alpha^*\right)\right)\right]} \tag{26}$$

To evaluate the parameters β_l and α_0^* we have fitted a smooth curve, by eye, to the experimental values of $1 - \tau/\tau_0'$, and used this curve to pick pairs of values (α^*, $\tau(\alpha^*)$) from which the quantity $0.11\beta_l(\alpha_0^* + \alpha^*)$ was calculated from equation (26). The quantity $0.11\ \beta_l\alpha_0^*$ was calculated from equation (25) by numerical integration, with the assumption $\tau_0 = \tau_0'$. When plotted against α^*, the values of $0.11\ \beta_l(\alpha_0^* + \alpha^*)$ calculated in this way lie nicely along a straight line, as required by the theory, for values of α^* in the range $0 \leq \alpha^* \leq 1000\ \mu c/kg$. Above $1000\ \mu c/kg$ the linear relation breaks down, reflecting the fact that the one experimental value in this higher range, at $2200\ \mu c/kg$, lies rather far from the theoretical curve. Ignoring this highest value we obtain in this way the parameters $\beta_l = 0.7 \times 10^{-7}\ day^{-2}\ (\mu c/kg)^{-1}$ and $\alpha_0^* = 200\ \mu c/kg$, from which the theoretical curve shown in Figure 3 is calculated. In addition to the point $\tau(0)$, and the points $\tau(\alpha^*)$ calculated from equation (24) over the range of validity of the Gaussian approximation, we have calculated the slope of the theoretical curve at $\alpha^* = 0$ from the following formula, which can be derived from equation (24):

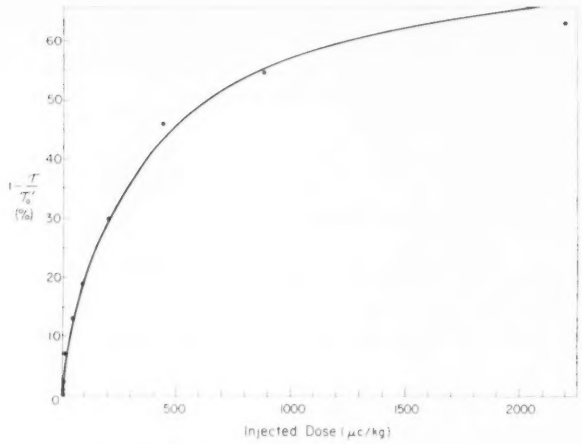

FIG. 3.—Percentage decrease in time to a 20 per cent incidence of bone tumors, as a function of injected dose of strontium-90. Solid curve is calculated from equation (26). Solid circles are experimental values reported by Dr. Finkel.

$$\frac{d}{d\alpha^*}\left(1-\frac{\tau}{\tau_0}\right)_{\alpha^*=0} = \frac{1-\dfrac{\beta}{10\beta_l}\dfrac{\displaystyle\int_0^{\tau_0} n^0(t)t^3dt}{\left[\displaystyle\int_0^{\tau_0} n^0(t)tdt\right]^2}}{10\gamma_0^2\tau_0^2 n^0(\tau_0)}\cdot\frac{1}{2}(0.11)\beta_l$$

where $\gamma_0 = {}^1/_2(0.11)\beta_l\alpha_0{}^*$. The ratio of integrals appearing in the second term of the numerator in this equation can be shown to have a value close to unity (actually 1.06).

A comparison of the parameters $\beta_l = 0.7 \times 10^{-7}$ and $\beta = 1.8 \times 10^{-7}$ suggests that of the radiation-induced deaths the fraction due to leukemia and related diseases in Dr. Finkel's experiments on mice is rather larger than has been estimated for man. A particular sensitivity to these diseases on the part of this strain of mice is suggested also by the large "background dose level" $\alpha_0{}^*$, reflecting the relatively large number of deaths due to these diseases in the control population.

Statistical Examination of Dr. Finkel's Conclusions.—In searching for evidence for the existence of a threshold body burden of strontium-90, below which no harmful effects are caused, Dr. Finkel uses two methods: (1) statistical analysis of the experimental data, and (2) extrapolation of experimental curves. We now consider these two methods.

The statistical analysis consists of a *t*-test of the hypothesis of no difference in response between the control population and a population dosed with strontium-90. Dr. Finkel accepts the null hypothesis at the 10 per cent significance level ("10% probability level or higher") for the three lowest-dosed experimental groups, and considers that this acceptance constitutes "evidence that there might be a threshold" or that "a threshold . . . may lie between 4.5 and 44 μc/kg."

It constitutes nothing of the kind. It is clear that the width of the acceptance

region for the null hypothesis (shaded region in Figs. 3 and 4 of Dr. Finkel's paper) should vary inversely as the square root of the number of animals in the experimental groups, assuming approximate normality of the death-rate curve $g(t)$, as Dr. Finkel must have done in applying the t-test. The threshold for which she finds "evidence" in the experiments is thus no threshold at all but simply a reflection of the statistical uncertainty of her information. It is clear that she could have found "evidence" of this sort for a threshold at any arbitrarily large radiation level (perhaps short of what would produce acute radiation sickness) by simply using few enough animals in her experiments.

The fallacy in Dr. Finkel's statistical argument is a failure to control the probability of type-II error of her test. Type-II error[5] is acceptance of the null hypothesis when it is in fact false. Consideration of the type-II error requires consideration of the alternative to the null hypothesis, which in this case is the theoretically likely linear response at low doses. If we use for the slope $d\Delta/d\alpha^*|_{\alpha^*=0}$ of the life-shortening response at low doses the value obtained above (eq. [13]) from a study of Dr. Finkel's results, namely, $d\Delta/d\alpha^* = 0.14\%\ (\mu c/kg)^{-1}$, and if we assume that Dr. Finkel's t-test acceptance region is appropriate to a one-tailed test at $\alpha^* = 8.9$ $\mu c/kg$, the highest experimental value for which the null hypothesis was accepted, then we can calculate the probability of type-II error. It is 85 per cent. This means that if there exists in fact no threshold at 8.9 $\mu c/kg$, Dr. Finkel's test would nevertheless have produced "evidence" for one in 85 experiments out of every 100 experiments performed. On the other hand, if there were in fact a threshold, the test would deny it in only 10 per cent of the experiments. Evidently the test is worthless as a proof of the existence of a threshold at this dose level (or lower, for which the probability of type-II error approaches the maximum that is possible, 90 per cent, for a 10 per cent probability of type-I error).

It is incumbent upon those who would extrapolate their threshold conclusions from 150 mice to 3×10^9 human beings that they demonstrate the existence of a significant experimental departure from the theoretically likely linear response, because although the existing burdens of strontium-90 are low, the number of individuals involved is very large, and the harmful consequences of proceeding on an unfounded assumption of a threshold are great.[6] As we have shown above, Dr. Finkel's results are in complete harmony with a linear law; in fact, the agreement between the linear law and the experimental results is better than could have been expected on the basis of the width of her null-hypothesis acceptance region.

As an alternative to the statistical tests, Dr. Finkel determines a threshold by extrapolating the experimental life-shortening curve. She states: "Since the [life-shortening] values for 1.3, 4.5, and 8.9 $\mu c/kg$ do lie along a straight line when plotted semilogarithmically, it may be argued that they represent true departures from the control value. An extension of this straight line crosses the control value at 0.4 $\mu c/kg$." It is difficult to see why the semilog plot rather than some other should be used for the extrapolation. But in fact an extrapolation of any kind is groundless. The three response values lie within less than half the range of probable error (within $-1/2$ P.E. to $+1/2$ P.E.) of the difference d between experimental values of Δ, as determined from the width of the null-hypothesis acceptance region ($\sigma_d = 5.5$ per cent). If it is not obvious that no non-zero regression slope determined from these points can have any statistical significance, one can show[7] that

the standard deviation of the regression slope estimator derived from the semilogarithmic plot is 2.2 times the estimated slope itself. If the semilogarithmically linear relation of the three points can be ascribed to anything but chance, then all of Dr. Finkel's statistical arguments are false. It is hard to imagine how two such mutually contradictory "proofs" could be advanced at one time.

There are other statistical points in Dr. Finkel's paper that merit scrutiny. In our discussion of her results we have had to rely on the correctness of the null-hypothesis acceptance region that she presents, but there are serious reasons for doubting its correctness. The width of the acceptance region corresponds to the estimate $\hat{\sigma} = 284$ days for the standard deviation of the death-rate function $g(t)$, if it applies to a one-tailed test on the difference between the life-shortening values obtained from two experimental groups of 150 animals each. On the other hand, Dr. Finkel's statement[8] that groups of 1393 animals would have been required to establish as significant at the 1 per cent level the difference observed (2.5 per cent) at the lowest dose corresponds to $\hat{\sigma} = 171$ days, a gross discrepancy. The latter value, we note, agrees reasonably with the values $\sigma = 161$ days from the Gompertz relation or $\hat{\sigma} = 191$ days from the agreement between experimental data and our theoretical curve.

It is clear that since the number of animals differs from one experimental group to another in Dr. Finkel's experiments, the null-hypothesis acceptance region cannot have width independent of injected dose α^*, as shown in her figures. From the information given there is no way to tell to which experimental groups the test appropriately applies.

More serious is the evident fact that Dr. Finkel applies the same acceptance region indiscriminately to the three very different sets of experimental data represented by her curves A, B, and C. It seems likely that the test was designed to handle the life-shortening data (curve A), because a t-test would not be inappropriate to life-span data, since the death-rate function $g(t)$ is (rather crudely) Gaussian. A statistical analysis of the curve-C data would be difficult, because the experimental statistic τ (20 per cent incidence time) is cumbersome to handle mathematically, as is evident in our discussion. But it is easy to show that Dr. Finkel's acceptance region is entirely inapplicable to the curve-B data ("proportion of animals that survived the latent period of 150 days and then died with osteogenic sarcomas").

The number of bone-cancer deaths in populations of a given size during a given time interval will be Poisson-distributed, if we neglect variations in population size due to deaths during the first 150 days, which is legitimate, as can be seen from Table 1 or from numbers given by Dr. Finkel, which show that the control group still contained close to 150 animals at $t = 150$ days. Whatever the low-dose regression function for curve B, it is clear from Dr. Finkel's Figure 4 that the expected number ξ of bone-cancer deaths is close to 3 for groups of 150 animals not dosed with strontium-90. To find acceptance regions for the null hypothesis of no significant difference in the number of such deaths between the control population and a dosed population of equal size we therefore find the value of the difference δ_P such that the probability of type-I error (one-tailed test) is P:

$$1 - P = e^{-2\xi} \sum_{\delta=-\infty}^{\delta_P} \xi^{\delta} \sum_{n=-\infty}^{+\infty} \frac{\xi^{2n}}{n!(n+\delta)!} \tag{27}$$

The results of a numerical evaluation of equation (27), in case $\xi = 3$, are $P = 0.10$, $\delta_P = 2.6$; $P = 0.01$, $\delta_P = 5$; $P = 0.001$, $\delta_P \cong 8.5$. Using the fact that the observed number of control deaths was 3 (2 per cent of 150), we find that the upper limit of the acceptance region at 10 per cent significance level is 3.7 per cent of 150. From Dr. Finkel's Figure 4 we therefore see that the highest strontium-90 dose that produced a "statistically non-significant increase" in the number of bone-cancer deaths is 8.9 μc/kg, not 200 μc/kg as stated by her. Her acceptance region represents for curve B a test having 0.1 per cent probability of type-I error. Her entire discussion of the statistical significance of the curve-B data is erroneous.

The Proper Testing of Evidence for a Threshold.—From the above discussion it is clear that a valid statistical test of the null hypothesis of "no response" at a given radiation level or a given dose of strontium-90 must use the type-II error as the basic parameter, rather than the type-I error, as is employed in the standard "cook book" tests, which are designed basically for application to the manufacture of goods. Alternatively stated, the null hypothesis that must be tested in the standard way is the hypothesis that the observed response values are in accord with a linear response curve at low doses.

Our analysis of Dr. Finkel's data for mice enables us to estimate reliably the linear decrease in life expectancy for low doses of strontium-90, and it is therefore possible for us to determine how many animals would have to be used in an experiment in which the mean lifetime of a control group is compared with the mean lifetime of a dosed group in order to establish the existence of a threshold at or above the dose used. We may consider two types of test: (A) the "minimal" test, that is, the test that requires as few animals as possible; (B) the "most powerful" test, which minimizes the probability both of type-I and of type-II error.

The null-hypothesis "no response" for test A is to be accepted if the dosed group exhibits no decrease in life expectancy, or an actual increase, when compared to the control group. Clearly this acceptance region makes the test minimal, because the probability of type-I error is 50 per cent, so that the test gives a neutral decision in case a threshold actually exists. If the number of animals used is greater than required for test A, then a decision as to the existence of a threshold will be more often right than wrong, in case that the threshold does actually exist. At the same time we can protect ourselves adequately against the serious alternative possibility by suitably choosing the type-II error.

Since the expected decrease in life expectancy for the dosed group is $\alpha \cdot d\Delta/d\alpha\,|_0$, the type-I and type-II errors are simultaneously minimized, and made equal, by choosing as the upper limit of the acceptance region for test B a decrease in life expectancy of $^1/_2\alpha \cdot d\Delta/d\alpha\,|_0 = 0.63\% \cdot \alpha$, where α is given in μc/kg retained in the body.

Since the expected decrease is proportional to α, the number of animals M required for the control group, if an equal number is used for the dosed group, is given by

$$M = \frac{\nu}{\alpha^2} \tag{28}$$

where ν is a constant that depends on the type of test (A or B), on the probability of type-II error, and on the standard deviation σ of the natural death-rate function

$g(t)$. Because M proves in all cases of interest to be large, it is adequate to use the normal distribution in computing the constants ν in equation (28), owing to the Central Limit Theorem.

TABLE 2

VALUES OF THE CONSTANT ν IN EQUATION (28)

	Type of Test			
	A ("Minimal")		B ("Most Powerful")	
σ (days)	284	170	284	170
10 per cent probability of type-II error	4560	1630	18240	5520
1 per cent probability of type-II error	14950	5360	59800	21440

Using all of these principles, we have computed the coefficients ν for the various circumstances shown in Table 2. In particular we compare the results for σ = 284 days, derived from Dr. Finkel's acceptance region, with the results for σ = 170 days, which seems most reasonable on the basis of the previous discussion.

The numbers ν given in Table 2 are equal to the number of animals in the control and in the dosed groups required to establish the existence of a threshold at α^* = 9.1 μc/kg, just above the highest injected dose (8.9 μc/kg) for which Dr. Finkel accepted the hypothesis that a threshold exists. The numbers of animals used in her experiments were too small by factors of 10 to 400, for the conclusion that she reached. By solving equation (28) for α we may compute very simply the lowest threshold $\alpha_T{}^*$ that could have been recognized with statistical significance in her experiments, assuming that 150 animals were used both in the control group and in the dosed groups, which in general was not the case (fewer were used). These values of $\alpha_T{}^*$ are 91 and 54 for test A and 181 and 109 for test B (in each case for σ = 284 days and 170 days, respectively). It is clear from the experimental data that no threshold exists at any of these levels, and accordingly we are required to conclude that Dr. Finkel's data show that there is no threshold large enough to have been recognized with statistical significance from her data.

Conclusions about Effects on Man of Strontium-90 from Fallout.—We now turn to the discussion of Dr. Finkel's conclusion that the present contamination with strontium-90 from fallout is so very much lower than the "threshold" levels that it is extremely unlikely to induce even one bone tumor or one case of leukemia.

This statement by Dr. Finkel is shown by the argument given above to have no justification whatever from her experimental results, obtained with 150 mice or fewer in her control group and injected groups. We may ask how many mice would be needed in each group in order to permit Dr. Finkel's statement to be made with statistical significance (or to be shown to be false).

The present average body burden of strontium-90 in the world's population is about 0.0002 μc. per person. This corresponds, with Dr. Finkel's conversion factor (5 to 10 μc. per 70-kg man equivalent to 1 μc. retained per kg for mice) to a retained dose α = 0.00002 to 0.00004 μc/kg in mice. Hence in order to justify Dr. Finkel's statement evidence would be needed that the mouse threshold is as great as about 0.00004 μc/kg; that is, we must place α in equation (28) equal to 0.00004 μc/kg. From the values of the constant ν in Table 2 (we use the values for σ = 170 days, which we believe to be better than those for σ = 284 days) we find $M = 1 \times 10^{12}$ for the "minimal" test and 3.4×10^{12} for the "most powerful" test with 10 per cent type-II error, and 3.3×10^{12} and 13.5×10^{12}, respectively, with 1 per cent type-II

error. We hence conclude that a study like that made by Dr. Finkel would have to use a much greater number of mice than the number of people in the world, in order to provide evidence that would justify her extreme statement to be made with statistical significance.

This conclusion is, of course, not at all unexpected. The difficulty of detecting by statistical methods an effect that causes a small increase in the annual number of deaths among the world's population is well known. For example, let us assume that the average number of deaths per year is 50 million. The statistical fluctuations from this average from year to year are measured roughly by the square root of this number, 7000; and accordingly a study of a larger population would be needed to show with statistical significance the existence of an effect resulting in an additional 1000 deaths per year (the rough estimate of the world-wide effect of the present body burden of strontium-90 from fallout, if there is no threshold). The same number of mice would be needed to test the equivalent effect in mice.

Summary.—We have developed methods of theoretical analysis of the results of experimental studies of the effects of injection of radioactive substances into animals on their life expectancy and on the incidence of tumors. These methods have been applied to the data reported for mice by Dr. Miriam P. Finkel, and it has been shown that her conclusion from these data that it is extremely unlikely that the strontium-90 from the fallout from nuclear weapons tests will induce even one bone tumor or one case of leukemia in human beings is completely unjustified.

* This paper is a contribution from the Division of the Geological Sciences (No. 908) and the Division of Chemistry and Chemical Engineering (No. 2421) of the Institute. A brief account of the work has been published by us (Letter to the Editor, *The New York Times*, Nov. 16, 1958), and a reply has been made by Dr. Finkel (*ibid.*, Nov., 30 1958).

[1] M. P. Finkel, *Science*, **128**, 637, 1958.

[2] E. B. Lewis, *Science*, **125**, 965, 1957.

[3] H. Jones, *The Nature of Radioactive Fallout and Its Effects on Man* (Washington, D. C.: Government Printing Office, 1957), p. 1109.

[4] Jahnke and Emde, *Tables of Functions* (New York: Dover Publications, 1945), pp. 1–9.

[5] A. M. Mood, *Introduction to the Theory of Statistics* (New York: McGraw-Hill Book Co. Inc., 1950), p. 247.

[6] L. Pauling, *No More War!* (Dodd-Mead & Co., 1958).

[7] Mood, *op. cit.*, pp. 292–294.

[8] Finkel, *op. cit.*, pp. 638–639.

*GRAVITATIONAL PROPERTIES OF ANTIMATTER**

BY L. I. SCHIFF

DEPARTMENT OF PHYSICS AND INSTITUTE OF THEORETICAL PHYSICS, STANFORD UNIVERSITY, STANFORD, CALIFORNIA

Read before The Academy, November 7, 1958

Eötvös Experiments and Equivalence Principle.—A very precise series of experiments involving gravity was performed by Eötvös and collaborators between 1890 and 1922.[1] These experiments sought for possible variations in the ratio of gravitational to inertial mass from one substance to another. By "gravitational mass"

we mean here what Bondi[2] refers to as the "passive gravitational mass," that is, a quantity proportional to the force exerted on a body by an external gravitational field, say of the earth or the sun. In contrast, one could also work with the active gravitational mass, which is proportional to the force exerted by a gravitating body; we shall return to a discussion of this quantity in the last section. By "inertial mass" we mean the ratio of some applied mass-independent force (electromagnetic, for example) to the acceleration of a body. We choose units such that th ratio measured by Eötvös is equal to 1 for normal matter.

The experiments were carried out for a number of substances, with varying degrees of accuracy. Three of these substances are especially convenient for our purpose, since they are nearly elemental and spread throughout the periodic table, and they are among those for which the ratio was measured with greatest precision: magnalium (90 per cent aluminum, 10 per cent magnesium), copper, and platinum. Comparison of magnalium and platinum showed that the difference in the ratio of gravitational to inertial mass was $(0.4 \pm 0.1) \times 10^{-8}$; this ratio difference was $(0.4 \pm 0.2) \times 10^{-8}$ for copper and platinum. The authors did not regard these small positive effects as significant, and we shall follow this interpretation of their work. We can therefore say that for these three substances there is no difference in the ratio to within about $^1/_2$ part in 10^8.

The assumption that there is no difference whatever in the ratio of gravitational to inertial mass from one substance to another leads to the equivalence principle and is basic to Einstein's general relativistic theory of gravitation. However, it is important to realize that the experiments that have been performed, while of very great accuracy, are not perfectly accurate, so that it cannot be said that the equivalence principle has been verified experimentally. This point has been emphasized by Dicke,[3] who noted that the energies involved in beta decay and in the self-gravitational interaction of a body of laboratory size would, even if they were to violate the equivalence principle, be far too small to affect the results obtained by Eötvös. In any event, we shall accept here the validity of the special theory of relativity, in accordance with which inertial mass and energy are interchangeable.

The Eötvös experiments are sufficiently accurate to show that the main factors that contribute to the inertial mass of a body also contribute equally or nearly equally to its gravitational mass.[3, 4] To this end, we first quote formulas for the amount $M\delta A$ by which the inertial mass would exceed the gravitational mass of an atom if one or another of these factors were to make no contribution to the gravitational mass. In this discussion, Z and A are the charge and mass numbers of the atom under consideration, which in practice is electrically neutral; δA is the difference between inertial and gravitational mass numbers; e and m are the electron charge and rest mass; M is the atomic mass unit (931×10^6 electron-volt$/c^2$); R is the nuclear radius; and c and $\hbar$ are the speed of light and Planck's constant divided by 2π. The pertinent formulas are as follows:

Rest masses of the atomic electrons:

$$M\delta A = mZ. \tag{1}$$

The kinetic energy of the electrons combines with the electromagnetic energy of interaction between electrons and nucleus to give the atomic binding energy; for Z not too small,[5, 6]

$$M\delta A \approx -15.6\, Z^{7/3} \text{ electron-volt}/c^2. \tag{2}$$

Electrostatic energy of the nucleus:[7]

$$M\delta A \approx 0.6Z(Z-1)e^2/R. \tag{3}$$

The kinetic energy of the nucleons combines with the mesonic interaction energies and the electrostatic energy (3) to give the nuclear binding energy:[8]

$$M\delta A = -B(Z, A). \tag{4}$$

In making a comparison between these formulas and experiment, we obtain sufficient accuracy by using the Z and A values for the most abundant isotope present in each of the substances mentioned above: magnalium, copper, and platinum. Table 1 presents the values of $|\delta A/A|$, in units of 10^{-8}, computed from equations (1)–(4) for each of these isotopes. The nuclear radius R is taken to be[9]

$$R = 1.24A^{1/3} \times 10^{-13} \text{ cm}. \tag{5}$$

TABLE 1

VALUES OF Z, A, AND $\delta A/A$ COMPUTED FROM EQUATIONS (1)–(4), FOR MOST ABUNDANT ISOTOPE PRESENT IN EACH OF THE SUBSTANCES INDICATED

			$\lvert\delta A/A\rvert \times 10^8$ from Equation			
SUBSTANCE	Z	A	(1)	(2)	(3)	(4)
Magnalium	13	27	26,400	25	144,000	890,000
Copper	29	63	25,200	68	242,000	923,000
Platinum	78	195	21,900	223	397,000	838,000

It is apparent from Table I that the magnalium-platinum and copper-platinum differences in each column are much greater than the experimental limit of $^1/_2$. Indeed, if we ignore the very unlikely possibility of fortuitous cancellations, we can conclude that the gravitational mass of an electron is the same as its inertial mass within a part in 10^4; that the atomic binding energy makes the same contribution to both masses within $^1/_4$ per cent; that the electrostatic repulsion between the nuclear protons has the weight expected from the equivalence principle to within two parts in 10^6; and that the nuclear binding energy decreases the weight of an atom by the expected amount to within a part in 10^5.

The same kind of argument can be used to show that positrons have the same or nearly the same gravitational mass as electrons and with the same sign. This demonstration is of some current interest because of suggestions by several physicists, mainly in connection with cosmology, that antimatter may have the opposite sign of gravitational mass from ordinary matter.[10] By "antimatter" is meant a substance in which normal particles (electrons, protons, neutrons) are replaced by antiparticles (positrons, antiprotons, antineutrons). Real positrons are not present in normal matter, so that it is necessary to make use of the virtual positrons that are produced because of the polarization of the vacuum by the Coulomb field of an atom. Since the probability amplitude for the presence of a virtual electron-positron pair is of order $Ze^2/\hbar c = Z/137$ (for Z not too large), we expect the mass difference to be

$$M\delta A \approx m(Z/137)^2 \tag{6}$$

if positrons have negative gravitational rest mass. This mass difference is calcu-

lated quantitatively in the next section; it is shown there that equation (6) has the correct order of magnitude and that comparison with the Eötvös experiments indicates that positrons have normal gravitational mass.[11]

A similar argument can be made in the case of antinucleons.[11, 12] Again, virtual antiprotons and antineutrons should be produced because of the polarization of the vacuum by the mesonic field of the nucleus. The corresponding $M\delta A$, if antinucleons possess anomalous gravitational masses, would be very much larger than that arising from virtual positrons, but it rests on a much less secure experimental and theoretical basis. The effect is expected to be larger because, first, the probability amplitude for the presence of a virtual nucleon-antinucleon pair should be of order unity rather than $Z/137$ and, second, the mass anomaly should be of order M rather than m. On the other hand, the theoretical structure of meson-nucleon field theory is not nearly so well understood as that of quantum electrodynamics, so that a reasonably reliable calculation is beyond the scope of present-day theory. Further, there is no direct experimental evidence for the existence of nucleonic vacuum polarization that is comparable with the evidence for electron-positron vacuum polarization. The latter, as produced by the atomic Coulomb field, manifests itself in the Lamb shift, where its contribution can be verified quantitatively.[13] For these reasons, we shall not attempt to compute the effect of an anomalous antinucleon gravitational mass on the Eötvös experiments, although it seems likely that the effect would be large; instead, we shall devote the next section to a quantitative discussion of the positron case.

We conclude this section by noting that several physicists[14, 15] have remarked that a violation of the equivalence principle leads, in general, to a violation of the principle of energy (or inertial mass) conservation. Suppose, for example, that a reversible reaction (nuclear or other) can take place in which the aggregate of reactants A on one side has, say, a smaller ratio of gravitational to inertial mass than the aggregate of reactants B on the other side. We further assume, tentatively, that energy is conserved; then the inertial masses of A and B must be the same, since the reaction can occur. Now, if we permit the reaction $A \rightarrow B$ to take place at a higher altitude in the earth's gravitational field than the reaction $B \rightarrow A$, then a cyclic process could occur in which B falls and A rises, and energy would be steadily withdrawn from the system.

In the particular case in which A is a nucleon-antinucleon pair and B is a pair of photons, Morrison[15] argues that energy conservation can be maintained by requiring that the inertial mass (or energy) of B be increased at the lower altitude before it is possible for the reaction $B \rightarrow A$ to take place. This increase in inertial mass is assumed to be of such a magnitude that it just uses up the energy that would otherwise be withdrawn from the system. The argument thus implies either that the inertial mass of A is greater at lower than at higher altitudes or that the inertial mass of B is greater at higher than at lower altitudes. Rather than accept such a dependence of inertial mass on gravitational potential, this dependence necessarily being different for different substances, we prefer simply to place the energy-conservation principle on the same basis as the equivalence principle. Both are valid with very great accuracy, but there is no assurance at present in either case that this accuracy is perfect. The idea that energy may not be precisely conserved is not

a new one and is, indeed, central to the continuous-creation hypothesis in cosmology.[16]

Gravitational Mass of a Positron.—The system under consideration consists of an electrically neutral atom in interaction with the electron-positron pair field and the earth's gravitational field. In the absence of gravity, the Hamiltonian is

$$H_0 = H_A + H_P + H_C, \tag{7}$$

where H_A, H_P, and H_C are the Hamiltonians for the atom by itself, the pair field by itself, and the (primarily Coulomb) interaction between the atom and the pair field, respectively. For our purpose, the gravitational potential of the earth can be regarded as static and very nearly independent of the space co-ordinates; the scale of distance variation is of the order of the radius of the earth. We call this potential ϕ and choose its units such that the total energy (or rest mass times c^2) of the atom in its ground state is changed from the value E_0 far from gravitating bodies to $E_0(1 + \phi)$ in the terrestrial laboratory. With this choice, the gravitational and inertial energies (masses times c^2) of our system in a particular state are simply the coefficient of ϕ in the corresponding energy eigenvalue and the ϕ-independent part of this eigenvalue, respectively.

We consider here three cases, which correspond to different assumptions concerning the effect of gravity on a positron: (*a*) A positron behaves in the same way as an electron. (*b*) The total gravitational energy (rest energy and kinetic energy) of a positron is equal to and opposite in sign from that of an electron. (*c*) The gravitational rest mass of a positron is equal and opposite that of an electron, but its kinetic energy is acted on normally by a gravitational field.

Case *a* corresponds to the equivalence principle. In this case the Hamiltonian (7) is replaced by

$$H_a = H_0(1 + \phi), \tag{8}$$

where we assume that the variation of ϕ over the dimensions of the atom can be neglected. Then the eigenvalues of equation (8) are clearly equal to $(1 + \phi)$ times the eigenvalues of (7), and the gravitational and inertial masses of the atom are precisely equal. However, as is well known, the lowest eigenvalue E_0 of H_A is, shifted by a logarithmically divergent amount, owing to the presence of $H_P + H_C$ and this must be included as part of the renormalization of the atomic mass. This renormalization is, of course, the same for the gravitational and inertial masses, so that they remain equal.

The simplest Hamiltonian that corresponds to case *b* is

$$H_b = (H_A + H_C)(1 + \phi) + H_P + \phi H_P'; \tag{9}$$

here H_P' is so chosen that it assigns a positive energy to each occupied positive energy (electron) state, and a negative energy to each unoccupied negative energy (positron) state. In the momentum representation,

$$H_P = \sum^{(+)} E_k a_k^* a_k - \sum^{(-)} E_k a_k a_k^*, \tag{10}$$

where the two summations are over positive and negative energy states E_k, as indicated, and a_k and a_k^* are electron-destruction and -creation operators, respectively. The simplest choice for H_P' is

$$H_P' = \sum^{(+)} E_k a_k^* a_k + \sum^{(-)} E_k a_k a_k^*. \tag{11}$$

We are interested in those deviations of the eigenvalue of H_b from the eigenvalue E_0 of H_A that are of second order in H_C and first order in ϕ. A straightforward perturbation calculation shows that, to this approximation (Z not too large),

$$E_b \approx E_0(1 + \phi) - (1 + 2\phi)\sum E_P^{-1}|\langle P|H_C|0\rangle|^2, \tag{12}$$

where the summation is over all electron-positron pair states P, of total energy E_P, and the matrix element of H_C is between the vacuum and pair states.

As remarked above, this summation is logarithmically divergent. The ϕ-independent part of (12) is, of course, the renormalized total energy (inertial mass times c^2) of the atom, since H_a and H_b differ only in their ϕ-dependent parts. However, the difference between the inertial and gravitational energies (that, is, the ϕ-independent part of E_b less the coefficient of the ϕ-dependent part) is no longer zero, as in case a, but is now equal to

$$\sum E_P^{-1}|\langle P|H_C|0\rangle|^2 \tag{13}$$

and hence is logarithmically infinite. It cannot be renormalized away and can be interpreted only by limiting the momentum space available to the pairs in some way. We shall return to (13) in the discussion of case c.

The simplest Hamiltonian that corresponds to case c is

$$H_c = (H_A + H_C)(1 + \phi) + H_P + \phi H_P'' \tag{14}$$

where

$$H_P'' = H_P - 2mc^2\sum^{(-)} a_k a_k^*. \tag{15}$$

Again, a perturbation calculation shows that

$$E_c \approx E_0(1 + \phi) - (1 + \phi)\sum E_P^{-1}|\langle P|H_C|0\rangle^2 - 2mc^2\phi\sum E_P^{-2}|\langle P|H_C|0\rangle|^2. \tag{16}$$

The quantity $M\delta A$ of the preceding section, the amount by which the inertial mass exceeds the gravitational mass, is just the negative of the last term of (16), divided by $c^2\phi$:

$$M\delta A \approx 2m\sum E_P^{-2}|\langle P|H_C|0\rangle|^2; \tag{17}$$

this quantity is finite, and we now proceed to compute its magnitude.

We substitute the interaction Hamiltonian $H_C = e\int\psi^* V\psi d^3x$, where V is the atomic electrostatic potential and ψ is the electron field operator, in (17), and perform the spin sums. The result is

$$M\delta A \approx [2m/(2\pi)^6]\int d^3k_1\int d^3k_2|U(\boldsymbol{q})|^2\,[E_1E_2 - h^2c^2(k^2 + \mu^2 - q^2/4)]/[E_1E_2(E_1 + E_2)^2], \tag{18}$$

where $h\boldsymbol{k}_1$ and $h\boldsymbol{k}_2$ are the electron and positron momenta, $\mu = mc/h$, $E_{1,2} = hc(k_{1,2}^2 + \mu^2)^{1/2}$, $\boldsymbol{q} = \boldsymbol{k}_1 + \boldsymbol{k}_2$, $\boldsymbol{k} = {}^1/_2(\boldsymbol{k}_1 - \boldsymbol{k}_2)$, and U is the Fourier transform of eV. Expressed in terms of the form factor F of the atomic charge distribution:

$$U(\boldsymbol{q}) = -(4\pi Ze^2/q^2)F(\boldsymbol{q}). \tag{19}$$

Relation (19) between U and F is so chosen that F is equal to unity when q is somewhat larger than the reciprocal atomic radius and somewhat smaller than the reciprocal nuclear radius; it becomes zero when q goes appreciably outside this range.

It is convenient to change the variables of integration in (18) from $\boldsymbol{k}_1$, $\boldsymbol{k}_2$ to $\boldsymbol{q}$, x, y, where

$$x = (k_1^2 + k_2^2 + 2\mu^2)/q^2 = (2k^2/q^2) + {}^1/_2\alpha,$$
$$y = (k_1^2 - k_2^2)/q^2 = 2(\boldsymbol{q} \cdot \boldsymbol{k})/q^2, \tag{20}$$
$$\alpha = 1 + (4\mu^2/q^2).$$

The sixth variable is the azimuthal angle of $\boldsymbol{k}$ with respect to $\boldsymbol{q}$ as polar axis and can be integrated over at once to give a factor 2π. The x, y integrations are performed next, with the range of y from $-(2x - \alpha)^{1/2}$ to $+(2x - \alpha)^{1/2}$, and the range of x from ${}^1/_2\alpha$ to ∞. The integration over $\boldsymbol{q}$ involves the atomic form factor and is done last. With the help of (19) and (20), equation (18) can be written

$$M\delta A \approx (4m/3\pi^2)(Z/137)^2 \int_0^\infty |F(q)|^2 f(\alpha) dq/q, \tag{21}$$

$$f(\alpha) = {}^3/_4 \int_{1/2\alpha}^\infty dx \int_0^{(2x-\alpha)^{1/2}} dy \, \frac{(x^2 - y^2)^{1/2} + 1 - x}{(x^2 - y^2)^{1/2}\,[x + (x^2 - y^2)^{1/2}]}, \tag{22}$$

where a slight simplification has been achieved by assuming that $F(q)$ is spherically symmetric. The y-integration in (22) can be performed analytically and the result cast in the following form:

$$f(\alpha) = {}^3/_4 \int_{1/2\alpha}^\infty dx \, \frac{(x - 1)[2x^2 - 2x + \alpha - 2x(x^2 - 2x + \alpha)^{1/2}]}{x(2x - \alpha)^{1/2}\,(x^2 - 2x + \alpha)^{1/2}}. \tag{23}$$

We wish to compute $f(\alpha)$, given by (23), for the range from $\alpha = 1$ ($q = \infty$) to $\alpha = \infty$ ($q = 0$). It does not appear possible to evaluate the integral in closed form, but an excellent approximation over the entire range of α can be obtained. For large α, the first four terms of the asymptotic expansion of $f(\alpha)$ are

$$f(\alpha) \approx (9\pi/32\alpha^{1/2})\,[1 + (1/12\alpha) + (3/128\alpha^2) + (5/512\alpha^3)] \tag{24}$$

The integral (23) can be evaluated exactly for $\alpha = 1$, and it is found that $f(1) = 1$. Finally, the behavior of $f(\alpha)$ for α slightly greater than unity can be found by a rather complicated expansion procedure; it changes linearly with a slope of ${}^{-3}/_4$. Thus for α slightly greater than, or equal to, unity, we have

$$f(\alpha) \approx 1 - 3(\alpha - 1)/4. \tag{25}$$

The leading term of the asymptotic expansion (24) is plotted on log-log co-ordinates as the dashed line in Figure 1, and all four terms as the solid curve. The dotted curve in the upper left corner of the figure is computed from (25) and drawn to join the solid curve so as to agree with a numerically computed point at $\alpha = 2$. This combination gives an excellent representation of $f(\alpha)$. It can be fitted within the plotting accuracy by the following formula, which is chosen to have the convenient form of the asymptotic series:

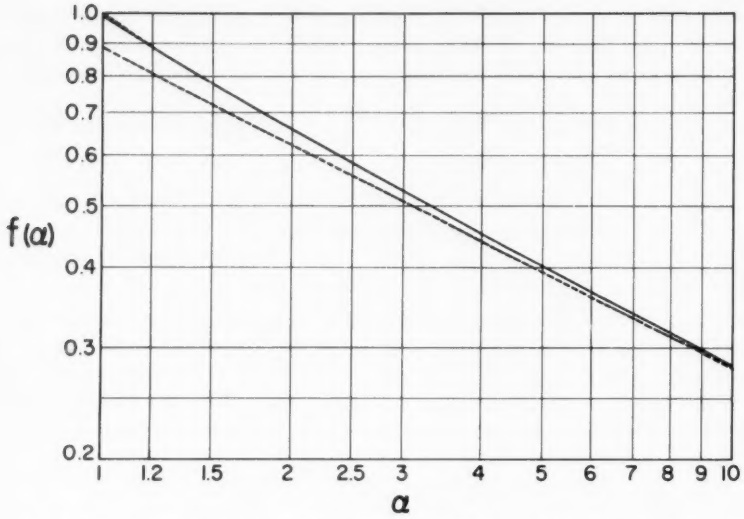

Fig. 1.—Plot of the leading term of eq. (24) (*dashed line*) and all four terms of eq. (24) (*solid curve*). The dotted curve in the upper left corner is computed from eq. (25) and is drawn so as to join the solid curve, in agreement with a numerically computed point at $\alpha = 2$.

$$f(\alpha) = \alpha^{-1/2}[0.884 + (0.118/\alpha) - (0.137/\alpha^2) + (0.135/\alpha^3)] \qquad (26)$$

The first term of (26) is the same as the first term of (24), and the other three coefficients are chosen to fit the value and slope at $\alpha = 1$ given by (25) and the numerically computed value at $\alpha = 2$.

Substitution of the leading term of (24) or (26) in (21) gives a good approximation for the mass difference:

$$M\delta A \approx (3m/8\pi)(Z/137)^2 \int_0^\infty |F(q)|^2 dq/(q^2 + 4\mu^2)^{1/2}, \qquad (27)$$

which was quoted as equation (1) in note 11. It is apparent from Figure 1 that this underestimates the mass difference, but never by as much as 12 per cent.

More accurate results than those quoted in note 11 can be obtained by substituting (26) in (21) and choosing, for $F(q)$, the Fourier transform of a uniformly charged sphere of radius R, where R is given by equation (5). A simple expression for the integral in (21) is obtained if terms of order $(\mu R)^2$ and $1/\mu a$ are neglected in comparison with unity, where a is the atomic screening radius:

$$\int_0^\infty |F(q)|^2 f(\alpha) dq/q \approx \ln (1/\mu R) + 0.338. \qquad (28)$$

Substitution of (28) in (21), with appropriate numerical values for the parameters, gives $\delta A/A$ for magnalium, copper, and platinum equal to 12, 25, and 53, respectively, in units of 10^{-8}. These numbers are somewhat larger than those quoted earlier[11] (10, 20, and 43); this is partly because the leading term of equation (24)

underestimate $f(\alpha)$ and partly because the assumption made concerning $F(q)$ in note 11 underestimates the large-q Fourier components of the nuclear charge density.

Thus, on the basis of case c, we would expect fractional differences of the ratios of gravitational to inertial masses of 13, 28, and 41 parts in 10^8, respectively, for magnalium-copper, copper-platinum, and magnalium-platinum. Since the experiments show differences of less than $^1/_2$ part in 10^8, we can conclude that the gravitational rest mass of a positron is the same as that of an electron within about 1 per cent. Thus the principal result of note 11 is not altered by the more accurate calculation presented here. The remaining approximations, having to do with the neglect of $(\mu R)^2$ and $1/\mu a$ compared to unity, the simple choice for $F(q)$, and the neglect of higher powers of $Z/137$ in the treatment of the atomic Coulomb field, will not significantly modify our conclusion. Further, we note that, since $E_P \geqq 2mc^2$, comparison of equations (13) and (17) shows that, no matter how interpreted, case b will always yield a larger value of $M\delta A$ than will case c.

The foregoing results depend on the specific assumptions made in equations (9), (10), (11), (14), and (15) concerning the forms of H_b and H_c. While these forms are simple, they are by no means unique. For example, the term ϕH_C could be modified so as to alter its pair creation and destruction parts without changing its electron and positron scattering parts; this last requirement means that the atomic electrostatic binding energy included in equation (2) does not contribute to the difference between gravitational and inertial masses. Another modification would consist in adding pair creation and destruction terms to H_P' and H_P''; these would not alter the effect of a gravitational field on a free positron. It is possible that such pair terms could mask the assumed anomalous gravitational properties of a positron and thus vitiate the conclusions of the last paragraph.[17]

If we multiply the pair creation part of ϕH_C by some number β and the pair destruction part by $\bar{\beta}$, in order to keep the whole Hermitian, we readily find that equation (12) is changed to

$$E_b' \approx E_0\,(1+\phi) - [1 + 2\phi Re(\beta)]\sum E_P^{-1}|\langle P|H_C|0\rangle|^2, \tag{29}$$

and equation (16) is changed to

$$E_c' \approx E_0(1+\phi) - \{1 + \phi[2Re(\beta) - 1]\}\sum E_P^{-1}|\langle P|H_C|0\rangle|^2 - 2mc^2\phi\sum E_P^{-2}|\langle P|H_C|0\rangle|^2, \tag{30}$$

where $Re(\beta)$ denotes the real part of the number β. Comparison of equations (12) and (29) shows that the choice $Re(\beta) = {}^1/_2$ would completely mask the assumed positron gravitational anomaly in case b. On the other hand, the form of equation (30) shows that no such masking is possible in case c.

The second modification mentioned above—addition of pair creation and destruction terms to H_P' and H_P''—leads to a more complicated calculation; contributions of order ϕZ, as well as ϕZ^2, would be expected. However, it turns out that if these pair terms have the simple form of a sum of products of positive-energy creation operator and negative-energy destruction operator for the same momentum and spin, together with the Hermitian adjoint, there is no change in equation (12) or equation (16). This result is directly attributable to the assumption of no momentum transfer in the newly added pair processes. If the spatial variation of ϕ were

taken into account, momentum transfer would occur; then the dependence on the radius of the earth would make such terms small and ineffective in masking.

Other possibilities for masking of the assumed anomalous gravitational behavior of a positron have not been explored. However, it seems likely, on the basis of the examples investigated, that it could occur only by a fortuitous combination of the assumed positron anomaly and a very special modification at some other point in the theory. Considerations of simplicity, taken together with the work of this section, lead us to conclude that positrons are very likely to have normal gravitational properties.

Passive and Active Gravitational Mass.—In the last two sections we have dealt with the passive gravitational masses and the inertial masses of particles and antiparticles. We can conclude with a very high degree of certainty that the inertial masses of positrons and antinucleons are positive, simply because energy is required to create them and is released when they annihilate. Furthermore, the experiments on the creation and annihilation of real electron-positron pairs make it virtually certain that electric charge is conserved in these processes; then the deflection of positrons in magnetic fields also shows that they have positive inertial mass.

A somewhat less definite conclusion can be drawn from the preceding sections: that the passive gravitational masses of antiparticles are also positive. One may then ask whether it is possible to account for a separation of matter and antimatter on a cosmological scale by means of gravitation.[10] The only possibility lies in assuming that the active gravitational masses of antiparticles are negative. As Bondi has shown,[2] such behavior may be consistent with the equivalence principle and the general theory of relativity, if the inertial and passive gravitational masses are precisely equal. All test bodies will be attracted by particles, and all will be repelled by antiparticles. An anomalous situation then arises when the motion of a particle-antiparticle pair under their mutual gravitational interaction is considered; this case is discussed in some detail by Bondi. However, it is sufficient for our purpose to point out that an aggregation of antiparticles by themselves would behave in a normal but unstable manner, in complete analogy to an aggregation of electrically charged particles all of the same sign.

Thus cosmological separation of matter and antimatter by means of gravitation requires that the passive gravitational masses of at least some kinds of antiparticles be negative. Then the equivalence principle is invalid, and general relativity cannot be used. While the interpretation of the Eötvös experiments discussed earlier in this paper probably precludes this possibility, it is of some interest to explore it further to see where it leads. When objects of both signs of passive gravitational mass are interacting, the anomalous behavior discussed by Bondi can be avoided by assuming that Newton's third law of motion is valid; for positive inertial masses this is the same as assuming that the active and passive gravitational masses of each object are equal in magnitude and sign. Then if at least the most massive antiparticles have negative gravitational mass, matter and antimatter will tend to separate from each other.

An interesting consequence of this model is that gravitational experiments performed in a region of the universe in which all particles are replaced by their antiparticles (antilaboratory) will lead to results different from those obtained when the same experiments are performed in a normal region of the universe (laboratory).

This is because, as was shown in the first section, electromagnetic and nuclear binding energy have positive gravitational mass. Now photons, which give rise to electromagnetic energy, and π mesons, which give rise to nuclear binding energy, are their own antiparticles. It follows that even if positrons, antiprotons, and antineutrons were all to have negative gravitational mass, there would be an asymmetry between the situations in the laboratory and in the antilaboratory.

It may be possible to preserve complete symmetry between laboratory and antilaboratory, while still having gravitational repulsion between the two, by drastically modifying the concept of gravitational field or by adopting an action-at-a-distance theory of gravitation. We could postulate that all photons, particles, π mesons, and their kinetic and intereaction energies attract each other; that all photons, antiparticles, π mesons, and their kinetic and interaction energies attract each other; and that all particles repel all antiparticles.[18] It seems unnecessary to attempt to develop such a model, however, since the preceding sections strongly indicate that particles and antiparticles actually attract each other.

In conclusion, it seems very likely that all particles and antiparticles have positive inertial and passive gravitational masses and that the equivalence principle is valid to at least very great accuracy. The active gravitational masses of antiparticles are then unknown. However, gravitational separation of stable aggregations of matter and antimatter on a cosmological scale cannot then be achieved by assuming that they are negative, and the validity of Newton's third law would require that they be positive.

* Supported in part by the U. S. Air Force through the Air Force Office of Scientific Research.

[1] For a summary of these experiments see R. v. Eötvös, D. Pekár, and E. Fekete, *Ann. Physik*, **68**, 11, 1922.

[2] H. Bondi, *Revs. Mod. Phys.*, **29**, 423, 1957.

[3] R. H. Dicke, *Revs. Mod. Phys.*, **29**, 355, 1957.

[4] A. H. Wapstra and G. J. Nijgh (*Physica*, **21**, 796, 1955) made use of earlier and less accurate work of Eötvös and so were limited in the conclusions they could draw.

[5] E. U. Condon and G. H. Shortley, *Theory of Atomic Spectra* (Cambridge: Cambridge University Press, 1953), pp. 338–39.

[6] It should also be noted that agreement between Einstein's prediction and the experimental value for the gravitational deflection of light would show that electromagnetic energy by itself behaves in accordance with the equivalence principle. However, there seems to be a discrepancy of 10–20 per cent; for a review of the experiments see E. T. Whittaker, *History of the Theories of Aether and Electricity* (London: Thomas Nelson & Sons, Ltd., 1953), **2**, 180.

[7] J. M. Blatt and V. F. Weisskopf, *Theoretical Nuclear Physics* (New York: John Wiley & Sons, Inc., 1952), p. 216.

[8] *Ibid.*, pp. 4–5.

[9] R. Hofstadter, *Ann. Rev. Nuclear Sci.*, **7**, 309, 1957; A. E. S. Green, *Revs. Mod. Phys.*, **30**, 569, 1958. It is sufficient for our purpose to regard the nucleus as a uniformly charged sphere.

[10] P. Morrison and T. Gold, *Essays on Gravity* (New Boston, N.H.: Gravity Research Foundation, 1957), p. 45; P. Morrison, *Am. J. Physics.*, **26**, 358, 1958; D. Matz and F. A. Kaempffer, *Bull. Am. Phys. Soc.*, Ser. II, **3**, 317, 1958; also private communications from others.

[11] A brief account of this work has been published: L. I. Schiff, *Phys. Rev. Lett.*, **1**, 254, 1958.

[12] The possibility of using virtual pairs, in connection with the Eötvös experiments, to determine the gravitational mass of antiparticles appears to have been first suggested by P. Morrison, *op. cit.*, p. 367, where a qualitative discussion concerning antinucleons appears. A similar qualitative remark and the result of a quantitative computation on positrons are given by Schiff (*op. cit.*); this work was completed without knowledge of Morrison's suggestion. Subsequently, Morrison

informed the author (private communication) that he had independently obtained the order-of-magnitude result embodied in eq. (6).

[13] See, e.g., S. S. Schweber, H. A. Bethe, and F. de Hoffmann, *Mesons and Fields* (Evanston, Ill.: Row, Peterson & Co., 1955), **1**, 299.

[14] Private communications.

[15] Morrison, *op. cit.*, p. 368; see also Morrison and Gold, *op. cit.*

[16] For a review of this theory see H. Bondi, *Cosmology* (Cambridge: Cambridge University Press, 1952), chap. XII.

[17] This possibility was suggested by S. D. Drell (private communication).

[18] A similar suggestion has been made by Morrison and Gold (*op. cit.*).

---

*SPECIFIC INTRACELLULAR BINDING OF RUBIDIUM BY RAT DIAPHRAGM MUSCLE**

BY P. MENOZZI,† D. NORMAN,‡ A. POLLERI,§ G. LESTER, AND O. HECHTER

WORCESTER FOUNDATION FOR EXPERIMENTAL BIOLOGY, SHREWSBURY, MASSACHUSETTS

Communicated by Theodore Shedlovsky, November 15, 1958

Introduction.—How cells discriminate between K^+ and Na^+ to maintain an unequal distribution of ions, potassium being high inside the cell and sodium outside, is a fundamental problem in biology. Currently, it is widely accepted that unequal ion distribution in muscle and nerve is the resultant of the activity of the cell membrane, postulated to possess energy-requiring pumps which transport K^+ into and Na^+ out of the cell against electrochemical gradients.[1] To account for ionic selection by these "active-transport" processes, the existence of specific carriers which discriminate between sodium (or lithium) and potassium (or rubidium) ions is postulated. It has alternatively been proposed that the selective accumulation of potassium is not primarily due to the "membrane" but is the resultant of intracellular mechanisms, such as "fixed charges" of structural proteins,[2] a charged-ordered semirigid lattice structure[3] or organized "water lattices."[4]

In a preceding paper[5] it was shown that the selective accumulation of rubidium against an apparent concentration gradient in *Neurospora crassa* is not explicable in terms of a membrane-pump mechanism but appears to involve a specific cytoplasmic binding system. This paper likewise presents evidence that the cytoplasm of rat diaphragm muscle exhibits properties of selective ion binding. Specifically, it will be shown that (*a*) the preparation of the usual hemidiaphragm involves cutting each muscle fiber of the diaphragm at two ends; (*b*) associated with this, a characteristic permeability feature of the cell is markedly altered, so that ions which are excluded from intact fibers readily penetrate into the cut fibers of the hemidiaphragm preparation; (*c*) under these conditions Na^+ and Rb^+ both enter the cell water of the cut fibers, but only the latter ion is accumulated against an apparent concentration gradient; and (*d*) the accumulation system for rubidium, which appears to operate with potassium as well, probably cannot be explained in terms of a membrane-pump mechanism.

Materials and method.—Two kinds of diaphragm muscle preparations, from male rats (Sprague-Dawley strain) weighing 100–180 gm. fasted 18–24 hours, were em-

ployed in these studies. One preparation was the isolated sectioned hemidiaphragm,[6] widely employed to study insulin action in vitro, where the diaphragm muscle is cut from the ribs on one side and the connective tissue attachments to the central tendon on the other, to yield, after trimming, two rectangular muscle strips. Since the muscle fibers run from the rib to the central tendon, each muscle fiber in the hemidiaphragm is cut at both ends and in our studies is about 5–8 mm. long. The other preparation, consisting of relatively uninjured fibers, was obtained by removing the diaphragm together with a section of the rib cage and the central tendon, in a manner similar to that described by Kipnis and Cori,[7] except that we included a larger portion of the rib cage, to minimize the possibility of injuring fibers of the diaphragm muscle.

The muscle preparations were incubated in Krebs-Hensleit Ringer's phosphate buffer[8] at pH 7.4 (henceforth referred to as "normal Ringer's"), potassium-free Ringer's, where Rb^{-86} (as RbCl) replaced K^+, or in modifications of these solutions, which will be described. Unless otherwise stated, incubation was carried out at 37.5° (with shaking) in the absence of added glucose, with O_2 as gas phase. With sectioned hemidiaphragms the volume of incubation medium employed was generally 5 ml.; cage preparations were generally incubated in 50 ml. of medium.

Experimental.—Cutting of muscle fibers results in marked changes in the permeability characteristics to various ions and sugars. To localize the basis for these alterations, we have examined the penetration of ferrocyanide into hemidiaphragm and cage preparations by histological means. When hemidiaphragm or cage preparations are incubated for 1 hour in a Ringer's solution, modified by substituting sodium and potassium ferrocyanide for NaCl and KCl, subsequent histological examination shows marked differences in ferrocyanide distribution between these preparations.[9] In the hemidiaphragm it was observed that ferrocyanide had penetrated into every fiber from both cut ends, leaving a central zone which had not been penetrated; in cage preparations ferrocyanide did not penetrate into any of the fibers and was restricted exclusively to the intercellular ground substance.[10] Study of ferrocyanide penetration in hemidiaphragms cut at only one end (the tendon was left in place at the other end), incubated for varying intervals of time, revealed that the entry of this large anion occurred exclusively via the cut ends and increased progressively with time at a rate estimated to be 1.8 mm. per hour from each cut end, over a 2-hour period.

Mannitol and sucrose, which are excluded from intact fibers, likewise penetrate into the intracellular water of the cut fibers of the hemidiaphragms, presumably via the cut ends. Thus the "space" of mannitol and sucrose was found to be 0.71 and 0.50 ml/gm muscle (wet wt.), respectively, whereas the inulin "space" in hemidiaphragms is 0.26 ml./gm muscle. The latter value is similar to those obtained for the intercellular volume of rat diaphragm in vivo, as determined by mannitol,[11] raffinose,[7] and inulin,[12] as well as to our own results in the cage preparation, where the "space" for sucrose and mannitol is equivalent to 0.24–0.28 ml/gm muscle. It would thus appear that the fibers in a sectioned hemidiaphragm possess two cut ends which permit non-selective diffusion of molecules like sucrose and ions like ferrocyanide but still restrict the entry of a large molecule such as inulin.

Despite these findings, which indicate that the cell membrane is no longer a con-

tinuous envelope in the fibers of the hemidiaphragm, the property of "active transport" of ions is retained. While hemidiaphragms incubated in Ringer's solution lose intracellular potassium and gain sodium to a much greater extent than do cage preparations, an equilibration of these ions between cell water and the external medium does not occur in cut fibers; the intracellular concentration of sodium is maintained, to a variable degree, below that of the medium, while the intracellular concentration of potassium is higher than that of the medium.[13] Following incubation in normal Ringer's solution, intracellular sodium gain is roughly equivalent, on the average, to potassium loss. This apparent reciprocity between potassium accumulation and sodium exclusion could suggest a primary causal relationship between these processes, as has often been proposed. However, these processes can be dissociated. Table 1 shows that, when hemidiaphragms are incubated in Ringer's solution where the external Na^+ concentration is varied from 10 to 148 mM (substituting sucrose for NaCl), potassium loss is quite constant, independent of whether or not sodium is gained. When the external Na^+ concentration is

TABLE 1

NA GAIN AND K LOSS IN HEMIDIAPHRAGMS*

CONCENTRATION OF Na^+ IN MEDIUM (μM/ML)	NO. OF HEMIDIAPHRAGMS	INTRACELLULAR (μM/ML CELL WATER)		VOLUME OF DISTRIBUTION OF NA IN INTRACELLULAR WATER (PER CENT)
		K Loss†	Na Gain‡	
148	38	110 ± 2.9§	+81.5 ± 3.0§	66 ± 2.0§
50	6	108 ± 4.0	+17.5 ± 2.2	72 ± 6.9
25	6	101 ± 3.9	− 3.7 ± 1.4	66 ± 4.1
10	6	129 ± 2.0	−10.0 ± 1.0	100 ± 8.0

* Incubated at 37.5° for 1 hour in Ringer's solutions modified by substituting sucrose for the NaCl removed.
† Initial K was 170 ± 2.0 μM/ml cell water.
‡ Initial Na was 20 ± 1.1 μM/ml cell water.
§ Standard error of the mean.

varied from 25 to 148 mM, the mean intracellular concentration of sodium achieved is equivalent to equilibration of Na^+ in about 70 per cent of the cell water. The sodium-exclusion process operative in cut fibers is exceedingly variable among individual tissues exposed to similar conditions; the basis for this variability has not, as yet, been elucidated.

Sodium exclusion in cut fibers appears to be an "active" process, in that it is inhibited by dinitrophenol (0.2 mM) or low temperature and to this extent resembles a "sodium pump." It should be noted from Table 1, however, that if a "sodium pump" actually were operative, the activity of this pump, but not its capacity, must be assumed to vary directly with the Na^+ concentration of the medium. Alternatively, it is possible to explain sodium exclusion on the basis that a certain fraction of the cell water is not available for Na^+ and that sodium gain represents an equilibration process in an available fraction of the total cell water. On this latter view, the maintenance of cell water in a state unavailable to Na^+ would require energy; a mechanism for sodium exclusion along these lines has been suggested by Baird *et al.*[4] Our present data do not permit a definitive decision between these dissimilar alternatives.

The "active-transport" properties of cut fibers are clearly revealed when the uptake of rubidium by hemidiaphragms is examined. Table 2 shows that hemidia-

TABLE 2

THE ACCUMULATION OF RB IN HEMIDIAPHRAGMS*

Incubation Time (Min.)	Intracellular Rb (μM/Ml.)	Rb_i/Rb_o
15	23.5 ± 1.1†(21)	4.6
30	41.1 ± 0.2 (75)	8.1
60	62.4 ± 0.4(118)	12.3
120	65.0 ± 4.0 (17)	12.7

* Incubated at 37.5° in K^+-free Ringers containing 5.1 mM Rb^+.
† Standard error of the mean; the figures in parentheses refer to the number of determinations.

phragms incubated in potassium-free Ringer's containing Rb^+ (5 mM) accumulate rubidium against an apparent concentration gradient, so that, after 1 hour, the ratio of Rb_i/Rb_o is about 12:1. Rubidium accumulation can be inhibited by addition of K^+ to the Ringer's solution containing Rb^+, as shown in Table 3, and it appears that the inhibitory effect of external K^+ is competitive in nature. As with

TABLE 3

EFFECT OF EXTERNAL K^+ CONCENTRATION ON RB ACCUMULATION IN HEMIDIAPHRAGMS*

Medium Concentration			Intracellular Rb (μM/Ml. Cell Water)		
K^+ (μM/Ml)	Rb^+	Hemidiaphragms	Total	Excess over External Concentration	Per Cent Inhibition
0	5.1	38	62.4 ± 0.4†	57.3	. .
2.6	5.1	5	48.4 ± 1.9	43.3	24.5
5.1	5.1	19	38.0 ± 0.5	32.9	42.5
10.2	5.1	5	28.6 ± 0.7	23.5	59.0
20.4	5.1	5	21.0 ± 0.6	15.9	72.3
122.0	5.1	5	6.0 ± 0.1	0.9	98.4

* Incubated at 37.5° for 1 hour.
† Standard error of the mean.

other active-transport systems, rubidium accumulation in excess of the external concentration is no longer observed when tissues are incubated in the cold (at or near 0°) or with dinitrophenol (0.2 mM) at 37.5°.

The rubidium-accumulation system in cut fibers has been sufficiently characterized to permit some conclusions concerning its nature. Thus rubidium accumulation is not dependent upon a "sodium pump," which, by expelling Na^+ against a gradient, results in the entry of Rb^+ against a concentration gradient. Figure 1 shows the results obtained in individual tissues where rubidium accumulation is plotted against sodium entry, measured in terms of equilibration of sodium in the cell water. It can be seen that the accumulation of rubidium is not dependent upon Na^+ exclusion. Indeed, the data demonstrate a tendency for rubidium accumulation to increase as the effectiveness of the sodium-exclusion process decreases.

It might be possible to account for rubidium accumulation by assuming a membrane mechanism which specifically pumps Rb^+ (or K^+) into the cut fiber at a rate sufficient to maintain an intracellular concentration of free Rb^+, 8 to 15 times in excess of the external concentration, despite efflux from the cut ends as well as the membranes. On this basis, it might be expected that, as the number of cut ends

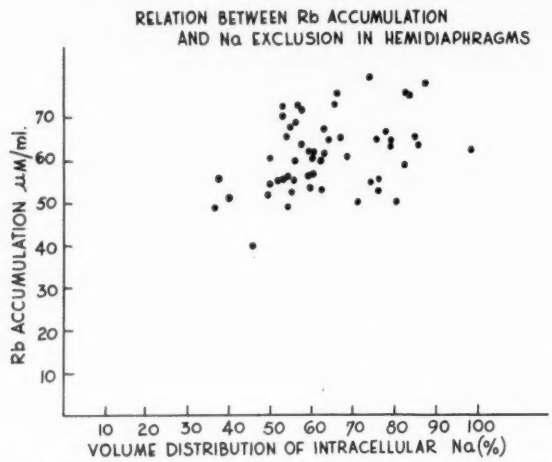

FIG. 1.—Each dot represents an individual experiment in a hemidiaphragm incubated for 1 hour at 37.5° in K^+-free Ringer's containing 5.1 mM Rb^+.

per fiber length are increased progressively (decreasing the ratio of the surface area of membrane to cut end), rubidium accumulation should decrease proportionately, since the cut ends may be regarded as offering no barrier to the efflux of free Rb^+. As shown in Table 4, when hemidiaphragms (2 cut ends) are further cut perpendicular to the fibers by a razor blade to give halves (4 cut ends) or thirds (6 cut ends), the expected proportional decrease in rubidum accumulation does not occur.

TABLE 4

EFFECT OF ADDITIONAL CUTTING OF HEMIDIAPHRAGMS UPON RB ACCUMULATION*

Preparation	Approx. Fiber Length (Mm.)	Intracellular Rb (μM/Ml. Cell Water)	Rb_i/Rb_o
"2 cut" (normal hemidiaphragm)	5–8	47.0 ± 1.71†(16)	9.2
"4 cut"	2.5–4	37.8 ± 2.77(8)	7.4
"6 cut"	1.7–2.7	30.4 ± 1.97(8)	6.0

* Incubated for 1 hour in K-free Ringer's containing 5.1 mM Rb.

† Standard error of the mean; the figures in parentheses refer to the number of observations.

These observations indicate that a Rb^+ pump mechanism is unlikely, unless one assumes that progressive cutting of the fiber results in an increase in the activity or efficiency of such a membrane pump.

If the accumulation and maintenance of high rubidium levels in the cell are actually the resultant of a rubidium pump, then factors which completely inhibit the uptake of rubidium in excess of the external concentration should have an equivalent effect on maintenance of previously accumulated rubidium. Nevertheless, as shown in Table 5, cold can prevent the initial accumulation of rubidium by hemidiaphragms, but, once rubidium has been accumulated intracellularly by prior incubation at 37.5°, it is retained intracellularly for several hours at 0°, so that the

TABLE 5

EFFECT OF COLD AND DINITROPHENOL ON ACCUMULATION AND MAINTENANCE OF INTRACELLULAR RB IN HEMIDIAPHRAGMS

Incubation Conditions	Intracellular Rb (μM/Ml)	Rb_i/Rb_o
Rb Ringer's* at* 37.5°		
60 min.	56.1 ± 2.2†(11)	11.0
120 min.	61.4 ± 6.1 (6)	12.1
Rb Ringer's at 0°‡		
60 min.	7.5 ± 0.3 (13)	1.5
120 min.	7.7 ± 0.5 (8)	1.5
Rb Ringer's + DNP (0.2mM) at 37.5°		
60 min.	8.1 ± 0.1 (5)	1.6
120 min.	7.9 ± 0.2 (6)	1.6
60 min. incubation at 37.5°	56.1 ± 2.2 (11)	11.0
Plus 5 min. at 0°‡	50.2 ± 1.9 (7)	9.9
Plus 15 min. at 0°‡	44.6 ± 1.1 (8)	8.8
Plus 60 min. at 0°‡	34.2 ± 0.8 (11)	6.7
Plus 120 min. at 0°‡	27.2 ± 2.0 (8)	5.3
60 min. incubation at 37.5°	56.1 ± 2.2 (11)	11.0
Plus 5 min. DNP (0.2 mM) at 37.5°	41.6 ± 1.5 (8)	8.2
Plus 15 min. DNP (0.2 mM) at 37.5°	23.4 ± 0.6 (10)	4.6
Plus 30 min. DNP (0.2 mM) at 37.5°	14.6 ± 0.7 (8)	2.9
Plus 60 min. DNP (0.2 mM) at 37.5°	11.4 ± 0.2 (9)	2.2

* 5.1 mM Rb^+.

† Standard error of the mean; the figures in parentheses refer to the number of determinations.

‡ Incubated in ice bath.

ratios of Rb_i/Rb_o are about 7:1 and 5:1 at 1 and 2 hours, respectively. Dinitrophenol (0.2 mM), on the other hand, causes a more rapid exit of previously accumulated rubidium than does exposure to cold. The dissociation of rubidium accumulation at low temperatures into separate energy-requiring processes for "formation" and "maintenance" excludes the possibility that a pump mechanism is responsible for rubidium accumulation in cut fibers.

The active-transport mechanisms which have been considered in relation to rubidium accumulation are based on the assumption that ions inside the cell are in the same "state" as in the external medium. Since these mechanisms do not adequately explain the observed findings with rubidium, some other mechanism must be involved. As an alternative, it may be proposed that rubidium accumulated in the cell in excess of the external concentration exists in an "associated" or "bound" state rather than as the free ion. Such a hypothesis, which emphasizes the im-

TABLE 6

EFFECT OF EXTERNAL Rb^+ CONCENTRATION UPON ACCUMULATION OF RB BY HEMIDIAPHRAGMS*

Medium Concentration (μM/Ml) Rb^+	Medium Concentration (μM/Ml) Na^+	No. of Hemidiaphragms	Intracellular Rb Total (μM/Ml. Cell Water)	Excess over External Concentration	Rb_i/Rb_o
2.55	150.5	6	48.2 ± 0.5†	45.6	18.9
5.1	148.0	18	62.4 ± 0.4	57.3	12.3
10.2	143.0	6	82.2 ± 3.0	72.0	8.1
25.5	127.5	6	98.8 ± 3.9	73.3	3.9
51.0	102.0	6	139.5 ± 3.1	88.5	2.7

* Incubated at 37.5° for 1 hour in K^+-free Ringer's.

† Standard error of the mean.

portance of cytoplasmic factors rather than the "membrane," would be consistent with the observed accumulation of rubidium by fibers which lack an intact cell membrane. The characteristics of the selective accumulation system for rubidium may be interpreted in terms of a selective cytoplasmic binding system. Table 6 shows that hemidiaphragms accumulate a relatively constant amount of rubidium in excess of the external concentration, when the Rb^+ concentration of the medium is varied from 10 to 51 mM. The binding system may thus be assumed to involve a limited number of sites (B), which can combine with Rb^+ to form RbB. The number of these sites may be quantitatively designated as the number of μM Rb per ml. cell water, in excess of the external concentration. In cut fibers there thus are about 70–90 sites per ml. cell water; whether or not all these sites are occupied depends on the external concentration of Rb^+. It may be assumed that B also interacts with K^+ to form KB, since the inhibition of rubidium accumulation by K^+ appears to be competitive. If K^+ and Rb^+ compete for sites, the question arises whether the rubidium accumulation observed in K^+-free Ringer's solution containing Rb^+ represents an exchange of external Rb^+ for intracellular potassium associated with B, or occupation of "vacant" sites, made available as a consequence of potassium loss from the injured fiber. Preliminary experiments indicate that rubidium accumulation occurs as a resultant of both processes.

The specific binding system is characterized by a marked sensitivity to low temperature and dinitrophenol, both of which completely inhibit the accumulation of rubidium; this suggests that the formation of RbB has an energy requirement. Maintenance of RbB likewise appears to have an energy requirement, since previously accumulated rubidium dissociates, to different degrees, when dinitrophenol is added or when the system is chilled to 0°. The fact that, at 0°, intracellular rubidium is maintained for 2 hours against an apparent gradient, whereas initial accumulation of rubidium is completely blocked at this temperature, indicates that the energy requirement for the maintenance of RbB is less than that required for its formation. The energy requirements of the binding system may be assumed to be relatively modest. The association of Rb^+ with B proceeds at 37.5° in the absence of an exogenous substrate, and addition of glucose to the medium has no significant influence; moreover, fluoride (1.0 mM) and cyanide (1.0 mM), both of which interfere with cellular metabolism, have little effect. Azide (1.0 mM), while possessing definite inhibitory action, does not abolish the association of rubidium with B as does dinitrophenol.

The accumulation of rubidium in rat hemidiaphragms has certain characteristics which are similar to those observed in *N. crassa*.[5] Thus, in both cases, it can be postulated that either Rb^+ or K^+ interacts with a specific binding system and there is competition between these ions for binding sites. The major difference between these two systems appears to relate to the stability of the RbB or KB, once it is formed. In *Neurospora*, RbB or KB does not appear to be readily dissociable or available for exchange; in diaphragm muscle, however, dissociation of RbB is readily effected by inhibitors such as dinitrophenol, azide, or, to a lesser extent, by lowering temperature. Moreover, as previously indicated, Rb^+ exchanges with a fraction of intracellular potassium which may be assumed to exist as KB. It is possible to explain these differences between muscle and *N. crassa* by assuming different types of B, as has been previously discussed.[5]

Our present information concerning the nature of the accumulation system in muscle is limited to the finding that it appears to be labile. Freezing and thawing or heating to 50° for a few minutes results in irreversible destruction of the system. Also, injury associated with cutting fibers reduces activity. There is insufficient information to permit profitable speculation upon the fundamental mechanism involved in the postulated intracellular binding system for Rb^+ or K^+. The limited data available could fit any hypothesis involving a metastable system permitting selective association of K^+ with binding sites, without necessitating the exclusion of Na^+. To this extent, the present data are consistent with either the theories of Ling[2] or those of Simon and Shaw.[3]

The results presented in this paper have an obvious bearing on the broad problem of the active transport of K^+ and Na^+. If the present findings on rat diaphragm muscle and the previous findings on *Neurospora*,[5] which indicate that most of the intracellular potassium may be selectively bound by a cytoplasmic system, should prove to be applicable to cells generally, certain widely accepted views concerning the role of the membrane in permeability processes may require basic revision.[14]

* Aided by grants from the National Science Foundation, Commonwealth Fund, and the Worcester Chapter of the Massachusetts Heart Association.

† Fellow of Institute of International Education, 1957–58: present address, Clinica Medica Generale, Università di Genova, Italy.

‡ Present address: Department of Physiology, Purdue University, Lafayette, Indiana.

§ On leave of absence from Clinica Medica Generale, Università di Genova, Italy.

[1] E. J. Conway, *Physiol. Rev.*, **37**, 84, 1957, and particularly the following, which deal with the proceedings of conferences relating to ion transport: "Active Transport and Secretion," *Symposia Soc. Exper. Biol.*, **8**, 1954 (New York: Academic Press); *Ion Transport across Membranes*, ed. H. T. Clarke (New York: Academic Press, 1954); *Metabolic Aspects of Transport across Cell Membranes*, ed. Q. R. Murphy (Madison: University of Wisconsin Press, 1957).

[2] G. N. Ling in *Phosphorus Metabolism*, ed. W. D. McElroy and B. Glass (Baltimore: Johns Hopkins Press, 1952), **2**, 748.

[3] S. E. Simon, F. H. Shaw, S. Bennett, and M. Muller, *J. Gen. Physiol.*, **40**, 753, 1957.

[4] S. L. Baird, G. Karreman, H. Mueller, and A. Szent-Györgyi, these PROCEEDINGS, **43**, 705, 1957.

[5] G. Lester and O. Hechter, these PROCEEDINGS **44**, 1141, 1958.

[6] C. L. Gemmill, *Bull. Johns Hopkins Hosp.*, **66**, 232, 1940.

[7] D. Kipnis and C. F. Cori, *J. Biol. Chem.*, **224**, 681, 1957.

[8] H. A. Krebs and K. Hensleit, *Z. physiol. Chem.*, **210**, 33, 1932.

[9] The incubations were carried out in Professor I. Gersh's laboratory at the University of Chicago by Dr. Joel Dennis and one of us. After incubation, the muscles were rapidly frozen, dried, and ferrocyanide visualized in 10-μ. sections as Prussian blue, according to the technic of I. Gersh and E. J. Steiglitz, (*Anat. Rec.*, **58**, 349, 1934). The sections were evaluated histologically by Drs. Dennis and Gersh. We should like to thank them for permission to include this material, which is part of a larger study to be published by them in a forthcoming issue of *Archives of Pathology*.

[10] After incubation of cage preparations for 2 hours, ferrocyanide penetrates a small number of fibers, indicating that, with prolonged incubation under our conditions, a deterioration of a cellular barrier to ferrocyanide occurs in some fibers.

[11] C. R. Park, L. H. Johnson, J. H. Wright, Jr., and H. Batsel, *Am. J. Physiol.*, **191**, 13, 1957.

[12] R. Creese, *Proc. Royal Soc. London, B*, **142**, 497, 1954.

[13] Tissue analysis of ions was determined as follows: the muscle preparations were removed (hemidiaphragms were cut from the cage preparation), blotted in a standardized manner, rapidly weighed, and the tissue extracted with boiling water. The concentrations of ions in tissue ex-

tracts and in the incubation medium were measured as follows: Rb^{86} was measured with a flow gas counter with suitable corrections for self-absorption; Na^+ and K^+ were determined with a flame photometer. The results for Rb, Na, and K are expressed in terms of μM per ml. cell water calculated on the basis of the following considerations. The total water content of both types of muscle preparations was found to be 0.80 ml/gm muscle. The inulin space in hemidiaphragms (0.26 ml/gm) and the sucrose space in the intact cage preparation (0.24 ml/gm) were taken to represent the intercellular volume in the two preparations, respectively, and were assumed to be in equilibrium with the external medium.

[14] After this paper has been submitted, E. Bozler, M. E. Calvin, and D. W. Watson (Am. J. Physiol., **195**, 38, 1958) reported studies which indicate that a large proportion of the intracellular K of the smooth muscle of frog stomach exists in a "bound" state.

SYMPOSIUM ON NEW DEVELOPMENTS IN THE STUDY OF COLOR VISION

By Invitation of the Committee on Arrangements for the Annual Meeting

Chairman: H. K. Hartline

Read before the Academy, April 28, 1958

CONTENTS

About Theories of Color Vision . . . . . *Yves LeGrand* 89
Studies of Color Blindness: A Unilaterally Dichromatic Subject . . . . . *C. H. Graham and Yun Hsia* 96
Color Vision: The Approach through Increment-Threshold Sensitivity . . . . . *W. S. Stiles* 100
Visual Pigments in the Intact Human Eye . . . . . *W. A. H. Rushton* 114
Color Vision in the Natural Image . . . . . *E. H. Land* 115

ABOUT THEORIES OF COLOR VISION

By Yves LeGrand

Museum National d'Histoire Naturelle, Paris, France

In this paper I shall try to give some kind of introduction to the following papers, which are dealing with recent experimental developments in the study of color vision. Color vision is a very old chapter of science, and it may be interesting to summarize briefly its main historical developments.

It seems that the French philosopher Descartes was the first, about the year 1630, to seek a physical explanation for colors. Before him, even those who thought that light was a modification of matter did not suppose that light might be endowed with any coloring property; they imagined color as a metaphysical quality of objects, a quality which could be perceived only when light fell on the object but was otherwise independent of light. Descartes understood that the color of an object is a change of the light after its reflection on the object, and he thought that this change might be related to the "spin" velocity of light particles.

On the basis of experiments on the mixing of pigments, several writers in the seventeenth century thought that all colors were a mixture of only three of them—say, red, yellow, and blue. This "trichromacy" was also derived by Mariotte from an examination of the solar spectrum, but he did not try to explain that curious fact. One century later, in 1756, a Russian physicist, Lomonosow, put forward the first theory of the three principal colors. He supposed there were three kinds of light particles, but this idea does not seem to have been widely accepted, and D'Alembert wrote in 1764 that there is a nearly infinite number of colors and not three, "as has been supposed by several physicists."

So trichromacy was in the air long before the beginning of the nineteenth century; but the great merit of Thomas Young was to establish in his celebrated 1802 paper that trichromacy was not a physical property of light but a physiological one of the

retina. The reason given by Young is very simple: he was convinced of the truth of Huygens' undulatory theory of light. In Newton's emission theory, it was quite simple to apply some form of "quantification" to the number of kinds of light particles, but it seemed impossible to do so at that time with the undulation theory, in which wave length is evidently a continuous variable. "As it is almost impossible," Young writes, "to conceive each sensitive point of the retina to contain an infinite number of particles, each capable of vibrating in perfect unison with every possible undulation, it becomes necessary to suppose the number limited, for instance, to the three principal colours, red, yellow and blue." As regards the choice of "principal colours," Young was, five years later, as a result of Wollaston's description of the spectrum, to modify this in favor of red, green, and violet.

Young's hypothesis was not followed up until it was revived by Helmholtz in 1852, and Helmholtz remarked that the choice of the "principal colours" remained arbitrary to a large extent. It was also under the influence of Young's ideas that Maxwell undertook in 1857 the measurements which gave to the trichromacy of vision an experimental foundation, though it was already implied in a rule given by Newton in 1704 for predicting the results of color mixtures. But Newton thought that his rule was not exactly valid, and, in fact, trichromacy was firmly established by Helmholtz and Maxwell a century ago.

The experimental fact of trichromacy may be expressed as follows: given any four lights, it is always possible to place three of them in one half of a photometric field and one in the other half, or else two in one half and two in the other, and, by adjusting the intensities of three of the four lights, to make the two halves of the field appear indistinguishable to the eye. This ability to match perfectly mixtures which are spectroscopically very different is a very striking fact in color vision.

These "matching experiments" are the basis of colorimetry. It is important to understand that in such measurements the observer is not asked what he is seeing but only which stimuli are equivalent to him: the eye is used only as a null instrument for establishing equalities. The mathematical form of any color theory restricted to the explanation of these matching experiments is very simple, owing to the algebraic character of the quantities used to define the stimuli and to the linearity of equations between them. These properties are known as "Grassmann's laws." The simple basis of all color theories obeying Grassmann's laws may be stated as follows: In the retina there are at least three different kind of receptors; any number greater than three, say n, is possible if we suppose that there may be $n-3$ additional linear relations between the n responses. The simplest possibility is evidently $n = 3$.

The first question is: What may be the physical basis for this spectral difference between retinal receptors? The most probable answer is that these differences may be due to various photosensitive pigments. Another possibility would be the existence of one single pigment but of different physical filters (colored oil droplets like those of birds and tortoises, or laminated structure of the outer limb producing interference filters); but the multiplicity of pigments is in accordance with the broad variety of visual substances for which there is evidence from studies of retinal extracts in different animals.

The second question is how to explain the $n-3$ linear relations between the receptors. It is easy to understand why a linear law governs the initial action of

light on any photosensitive substance, but linearity is lost as soon as a nervous message is sent from the cone to the bipolar cell; then it seems necessary to suppose that our $n-3$ relations are of a physical, not a physiological, kind. The simplest way of obtaining such relations by physics is the mixture of three different photosensitive pigments only; if the maximum absorption of each pigment remains low, the relations are very nearly linear.

It is easy to understand why all the experimental laws of matching experiments result immediately from the existence of three pigments, whatever the mixture made eventually with them in the retinal cones may be: equivalent stimuli are such that the three initial photochemical actions are identical, and anything which occurs afterward cannot alter this identity. Colorimetry may be affected neither by any mixture of pigments nor by nervous phenomena.

If three photopigments exist in the retina, it would be very important to determine their absorption curves. But, unhappily, matching experiments alone are inadequate to determine a unique set of 3 spectral sensitivities which would correspond to the spectral absorptions of the pigments, after correction for selective losses through the ocular media or the macular pigment. Owing always to linear Grassmann's laws, any trilinear transformation of such a set gives another set which is exactly equivalent from the colorimetric point of view. In practice, a very artificial but convenient set has been chosen by the CIE under the name "XYZ system." In this system of spectral sensitivity curves (Fig. 1) the $\bar{y}$ curve is identical with the photopic efficiency curve of cones; the $\bar{x}$ curve has two maxima and a minimum, almost zero, near 500 mμ; it does not indicate any particular characteristic of vision and comes from the arbitrary choice of this wave length for the nearest approach of the side YZ of the triangle of reference stimuli to the spectrum locus.

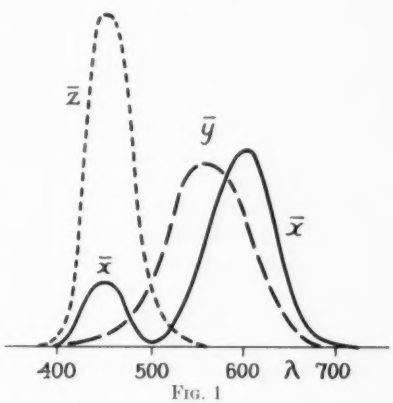

FIG. 1

In the past, a great number of attempts have been made to add supplementary conditions that might limit the infinite variety of possible spectral sensitivities of the retinal photopigments. These conditions were sought in a study of dichromatism, assuming, with Helmholtz and his pupil König, that the reduction from trichromatism to dichromatism results from the suppression of one kind of pigment. For the protanopic sort of dichromatism, this hypothesis is probably true, but for the deuteranopic one it is rather uncertain, and Fick's interpretation seems to be better: the reduction would take place by unification of the responses of two kinds of retinal receptors which are independent for a normal observer. If so, the study of dichromatism is not able to give sufficient information about the actual absorption curves of the photopigments, and the only ways to obtain these curves should be the extraction and the chemical study of the pigments, as Dr. Wald did in his celebrated work on iodopsin, or the physical study of light absorption by pigments *in situ*, as in Dr. Rushton's beautiful experiments.

We have said that the simplest way of explaining matching experiments was to suppose the existence of three photopigments, colorimetry being unaffected by any mixture of these pigments or by nervous interactions. But in theory it would also be possible to suppose a number of pigments larger than three, if supplementary conditions of linkage between the responses were assumed. Brindley has shown that this last hypothesis is in conflict with experience, by studying the breakdown of color matches in the case of intense adaptation. There is direct evidence, given by Rushton, that adapting lights, if bright enough, can bleach away a large fraction not only of the rhodopsin in the rods but also of the cone pigments of the retina. In fact, conspicuous changes in the appearance of colors can be produced by adaptation to colored lights without upsetting color matches between lights of different spectral composition, as had been already shown by Von Kries. With very bright adapting lights, color matches begin to be disturbed, probably owing to the self-screening of the photopigments and of their decomposition products. But in this case Brindley has shown the very remarkable fact that all the adapting lights, whatever their wave length, cause the same change for a given luminance in the amount of red 680 and green 550 necessary to match a yellow 580 (Fig. 2). This is a proof that there are not more than three pigments. In any other hypothesis it would be necessary to postulate compensation mechanisms, which are highly improbable from a physiological point of view.

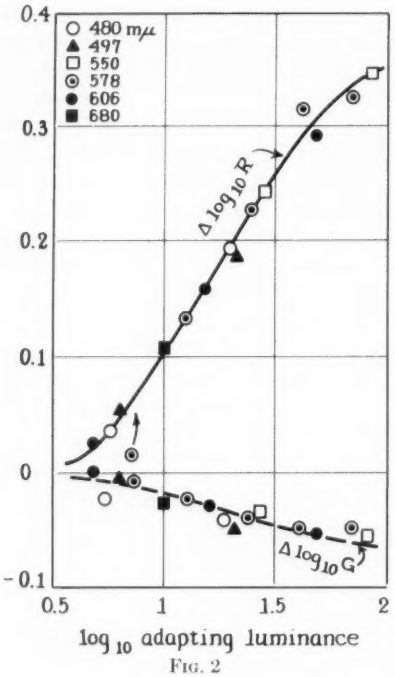

FIG. 2

If luminance has only a small effect on matching experiments in the fovea, it has a much larger one in the extrafoveal retina at low levels, because the scotopic efficiency curve of rods differs from the photopic one of cones (Purkinje shift). As Brindley has shown, this is a direct proof that rhodopsin is not one of the three pigments of color vision.

In conclusion, all work done in matching experiments and their breakdown at very high or very low levels of illumination is in good agreement with the very simple assumption of three, and only three, independent photopigments in the human retina. If $n = 3$, does it mean that there are also three types of cones in the human retina, each of them containing one of the three pigments? Not necessarily. In fact, apart from the regular change of form from the center of the fovea to the extreme periphery, all the cones seem to be alike, whereas one might expect to observe anatomical differences either causing or caused by the difference in the pigment con-

tained in the cones. There is another argument against the three types of cones, given by Abney's law of luminance additivity: it is not possible to suppose that certain bipolars or ganglion cells may add the impulses of the three types of cones and so respond to every wave length, because in these combined impulses any linear characteristic is lost. On the contrary, a fourth type of cone containing a mixture of the three pigments would give a very simple way of interpreting Abney's law; if this fourth type is the most frequent, it also explains the fact, known from Aubert onward, that a very small colored object on a neutral background of the same luminance appears colorless and, more generally, that color vision is fully developed with large field tests only. In the three-cones theory, on the contrary, the opposite effect should occur, and a white point should look colored in the neighborhood of the foveal threshold because it would excite only the cone with the lowest threshold.

If it seems necessary to suppose at least four types of cones, the simplest assumption about the mixing of the three pigments is that it is made at random. Then the majority of the cones would contain the three pigments mixed approximately in equal concentrations, and their spectral response would be roughly the photopic efficiency curve; a smaller number of cones would contain a relatively high concentration of one or two pigments, and it would be possible to classify them in six groups, although, in fact, their spectral sensitivity varies in a continuous manner. This hypothesis would seal the reconciliation between Young and Newton, between the sacred numbers 3 and 7, three pigments and, roughly speaking, seven kinds of cones in the retina!

As matching experiments are of no value in teaching us anything on this subject of retinal receptors, it is necessary to find another line of approach. The first may be given by electrophysiology, and the thresholds for spike discharges obtained by the microelectrode methods have led to Granit's familiar dominator and modulator curves. Such results are unhappily difficult to interpret, because they reflect neural interactions which modify the spectral responses.

A second kind of information is given by the size of color discrimination steps. Let us recall Dr. MacAdam's well-known ellipses at a constant level of luminance, which are of great practical value but rather difficult to interpret theoretically, because we do not know at what part of the visual pathway this form of discrimination is obtained. Dr. Stiles's results obtained by the two-color method are simpler, but here also the assumption made that the threshold of each mechanism is independent of the activity of the others may not be valid, perhaps, when different mechanisms are very differently active (such interactions between sound and light have been observed); nevertheless, Dr. Stiles's experiments give us the hint that n is at least 5.

A third line of approach is obtained through artificial dichromatism (and even monochromatism) produced by adaptation to very bright lights. The most interesting result is obtained through adaptation to intense yellow light because, besides the loss of color discrimination, there appears a severe defect of luminance discrimination. This seems to indicate that the "blue" pigment plays only a minor role in the luminance receptor where the three pigments are mixed. Another difference between the "blue" cones and the others is the low visual acuity associated with the blue mechanism, perhaps because synaptic convergence is greater for these receptors.

In conclusion, if colorimetry is very simply and completely explained by the existence of three photopigments, the nature and even the number of different retinal receptors is still unknown; but there are some reasons to think that this number is greater than three.

Finally, it remains to explain how colors look to us, and this psychological part of color theories is perhaps the most difficult to place on a physical and physiological ground. Young's trichromatic idea meets here with its two main difficulties; the first is the nature of yellow sensation, which does not at all appear as an addition of green and red; the second is the total neutralization of the color attribute of sensation, in complementary pairs, for example. Both necessitate some kind of inhibition, which may be found perhaps in the retina itself. We may imagine a compound receptor containing a mixture of green and red pigments, the excitation of which by yellow light would deliver a special message to the optic nerve; but it is necessary to suppose also that in "green and "red" receptors, the messages cancel one another when they are equal. An inhibition of the same kind would suppress the responses of red, green, and blue cones when they are in equilibrium and leave only the message of the white compound receptor, as in Pieron's theory of "receptor tetrad." All this is physiologically possible, but far from being proved.

More than eighty years ago, Hering proposed a general principle of retinal inhibitions, produced by opposite actions of light on three substances; it is difficult to find a physiological basis for such opposed responses given by one retinal mechanism; perhaps it might be looked for in the duality of the "on" and "off" responses that certain recent researches in electrophysiology ascribe to a single retinal receptor.

Another possibility is postulated in "zone" theories, such as those of Müller and Judd, which combine Young's peripheral trichromacy with more complicated structures at another stage of the visual pathway. The difficulty is that a large number of such theories may be constructed with any of these psychological data but that none of them can be proved as being the right one. It would be necessary to know more of the mechanism of vision in the cortex because experimental evidence obtained at this level, say by electrical records, has until now been rather poor.

Some fifteen years ago, LeGros Clark thought that the geniculate body might play an important part in color vision, its six layers being related with trichromacy, but now his ideas seem to be abandoned, or nearly so. In the visual cortex of the cat, Chang had registered electrically a complex response to the excitation of the eye by a flash of light; here again he had thought that this fact was related to trichromacy, but unhappily we are now almost certain by psychological experiments that the cat does not perceive colors.

In fact, the only thing which is quite sure is the extraordinary complication of the visual cortex; each optic nerve contains about one million fibers, while the visual cortex has near 1,400 million neurons. This multiplication is specially high for the foveal center, each fiber which comes from the fovea being connected with 5,000 cortical neurons, according to Sholl. Therefore, it is easy to imagine the complexity of interactions and inhibitions that may arise in so intricate a network of neural synapses. It is easy to imagine it, but it has been impossible until now to apply any cortical theory even to the simplest psychological laws of color vision.

For example, color contrast is perhaps explainable through retinal adaptive effects

at the line of separation between two areas the colors of which differ, but there exist also opposite phenomena of interaction between contiguous colors, which lead to a diminution of the chromatic difference instead of its exaggeration. Fuchs, who discovered this effect, named it "equalization," and it has recently been studied by the psychologist Musatti. Let us give an example of such an equalization effect: red circles which are actually identical in color everywhere seem more purple on a blue ground and more orange on a yellow one. Professor Kanizsa has described another similar effect, the action of the marginal gradient of the figure upon its chromatic appearance: the lack of abrupt separation between a red ground and yellow bars changes the red tonality of the ground which orange equalization makes less different from the yellow of the figure. Whether such phenomena are retinal or cortical or both, we do not know. Though they seem rather simple when we try to explain them by words as psychologists do, they disclose probably the astonishing complexity of synapses in the brain (or in retina—but the retina is a peripheral part of the brain).

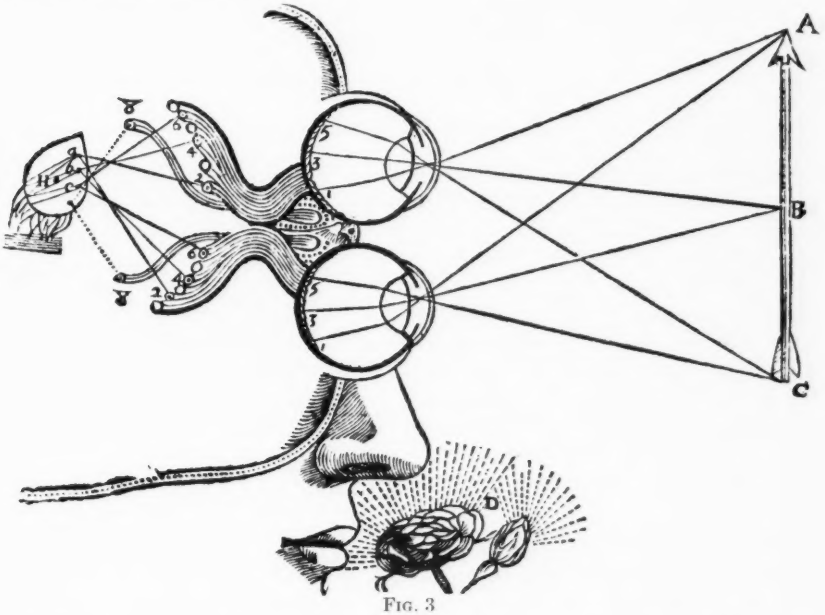

FIG. 3

And how can we explain color constancy, one of the major facts in the psychology of color vision, which allows us to identify objects in spite of the physical variations in the intensity and composition of the light which falls on them? It is an aspect of the great problem of generalization and abstraction, which was analyzed by Professor Hebb in terms of cortical connections in his book, *The Organization of Behavior;* by Pitts and McCulloch by formal mathematical methods; and by Beurle by means of a probability theory of the aggregates of interacting neurons. But I agree entirely with Sholl's conclusion that "at this time it would be premature to attempt to put forward a complete theory of cortical organization" which would

explain in particular the psychological laws of color vision. I expect that in the future anatomists and physiologists will add greatly to our knowledge in these very interesting problems, but as a conclusion I think best to remind you of an old picture in Descartes's book on man (Fig. 3). Our modern representations would not differ much from Descartes's, except that the "pineal gland," which he supposed to be the"seat of imagination and common sense," is now out of date and replaced by the cortex. As to the transmission of visual messages to the brain and storage of visual engrams in the cerebral substance, we have not added a great deal to this picture of the seventeenth century. It is sometimes good, for the modern scientist who is so proud of the progress of science, to learn modesty in an old French book which was written three hundred years before the beginning of this so-called "atomic age."

STUDIES OF COLOR BLINDNESS: A UNILATERALLY DICHROMATIC SUBJECT

By C. H. Graham and Yun Hsia

COLUMBIA UNIVERSITY, NEW YORK

Since the time of Thomas Young[1] considerations of color blindness have played an important and possibly primary role in determining the forms of various theories of color. The literature of color blindness reserves an important place for studies of unilaterally color-blind subjects, that is, persons who are color-blind in one eye but normal in the other. Judd[2] states that thirty-seven cases of unilateral color blindness have been reported in the literature, but, because of inadequate procedures, equipment, or background information, only about eight have proved useful for theory, and then not in all desirable details. No ordinary color-blind subject can tell us how the colors he sees compare with those seen by a normal subject, but this is precisely what a unilaterally color-blind subject can do. He can make a direct comparison of colors seen by his color-blind and normal eyes and thus provide us with a basis for relating the color perceptions of the color-blind eye to those of the normal.

We have recently had an opportunity to examine some visual discriminations of a unilaterally dichromatic subject. The results of that investigation form the content of this discussion.

Color Mixture.—The fact that our subject is unilaterally color-blind was established well enough in preliminary tests, but its precise specification was unequivocally established by her data on color mixture.

The color-mixture experiments show that our subject's normal eye can match any wave length of the spectrum by a mixture of three primaries—460, 530, and 650 mμ. In addition, they give data that are in accord with Wright's standard results in 10 normal subjects. The luminance units involved in the settings fall within the distribution given by Wright's subjects. In a word, our subject's right eye has normal color vision.[3]

The data on the color-blind eye are different from the data on the normal eye.

In her color-blind eye our subject can match any wave length of the spectrum with a combination of two primaries, 460 and 650 mμ. Judging by this performance, we can say that her left eye is unequivocally dichromatic. Her results are in accord with Pitt's standard data[4] on dichromats.

The color-mixture curves for our subject's left eye show that that eye is dichromatic, but they do not tell what type of dichromasy is represented. Two rare types of dichromasy are excluded, but the data do not differentiate between two forms—deuteranopia, the most frequently encountered form (sometimes inaccurately referred to as "green blindness"), and protanopia, the next most frequently encountered form (sometimes inaccurately referred to as "red blindness").

Spectral Sensitivity.—Differentiation between the two forms may be accomplished in a number of ways, all depending upon a testing of the sensitivities of the eye to red light. If a dichromatic eye is relatively insensitive to red light, it may be classified as protanopic. If the eye is sensitive to red and at the same time gives the same kind of color-mixture curve as our subject's dichromatic eye, then it is deuteranopic. Some data by Hsia and Graham[5] and by Hecht and Hsia[6] on groups of protanopic, deuteranopic, and normal subjects indicate that deuteranopes usually lose sensitivity in the blue and green regions of the spectrum.

The spectral sensitivity curves of our unilaterally color-blind subject,[7] shown in Figure 1, demonstrate that her dichromatic eye loses sensitivity in the blue-to-green region of the spectrum and has normal sensitivity in the red. She is therefore classifiable as deuteranopic. In addition, her data on spectral sensitivity strengthen the interpretation that deuteranopes usually show a loss of sensitivity in the blue-to-green region of the spectrum. Her loss in this region is well marked.

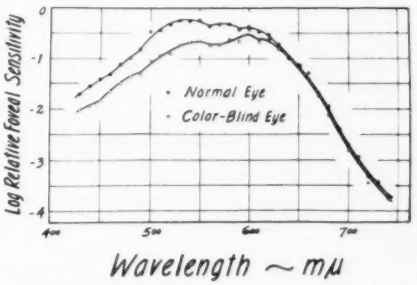

FIG. 1.—The spectral sensitivity curves for a unilaterally color-blind subject. The upper curve for the normal eye; the lower one, the curve for the color-blind eye. Data of Graham and Hsia (these PROCEEDINGS, **44**, 46–49, 1958).

Binocular Color Matches.—As part of our analysis of our subject's vision we performed an experiment on binocular color matching. This experiment had to do with the question: What colors does our subject see in her color-blind eye? The subject regarded the color of a vertical slit presented to her left eye in the visual field and simultaneously observed the color given by a horizontal slit in her normal eye. The wave length of light stimulating the normal eye could be changed until a narrow wave-length band was obtained that, in the opinion of the subject, matched the color seen in the dichromatic eye.

The results of the experiment are summarized in Figure 2. In general, it seems that in her dichromatic eye the subject matches all wave lengths greater than the neutral point (i.e., the narrow wave-length band that a dichromat sees as gray) with a wave length seen in the normal eye lying near 570 mμ. Wave lengths shorter than the neutral point in the dichromatic eye are matched in the normal

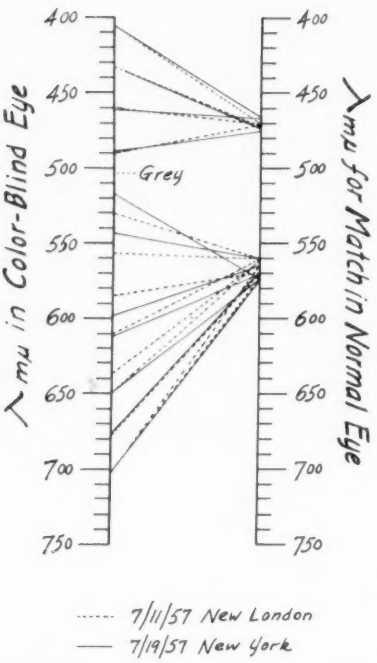

FIG. 2.—Results of the experiment on binocular matching

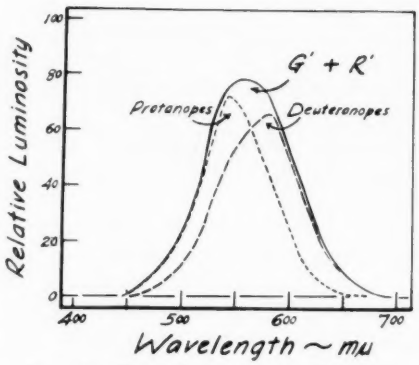

FIG. 3.—Theoretical curves. The curve drawn with a continuous line represents the sum of R' and G' at any ordinate value. R' comes about when the R curve is "transformed," i.e., shifted to shorter wave lengths (with change of shape); G' comes about when the G curve is shifted to the longer wave lengths. The ratio of G' to R' is unity and $R' + G'$ is equal to normal spectral sensitivity (except for the small contribution of a curve for blue). The right dashed-line curve represents the spectral sensitivity curve for deuteranopes, and the left dashed-line curve, the curve for protanopes.

eye by a blue of about 470 mμ. This result is in line with Judd's analysis[7] of earlier data.

Discussion.—Our subject shows a loss of spectral sensitivity in her color-blind eye and at the same time sees all wave lengths above 502 mμ as yellow. How can she (or, in fact, any deuteranope) see yellow (which we take to be a mixture of red and green) in the total or even partial absence of a green brightness mechanism?

To begin with, it seems that, in order to account for the seeing of yellow (we shall not deal with blue!) by deuteranopes, we may appeal to the Fick[8] and Leber[9] theory. This theory says that yellow is seen by deuteranopes because, in such persons, the red and green receptors become indistinguishable so far as sensory input is concerned. However, they still maintain their usual central connections, and thus, on stimulation by wave lengths longer than about 500 mμ, they provide a mixture of yellow due to excitation of the usual red and green centers. This latter consideration sets a requirement. (In discussing it, we shall refer to the normal red and green processes as R and G and their possible transformations, sensory or central, as R' and G'.) (1) It is required that spectral luminances for a mixture of yellow determined by R' and G' must be in nearly the same ratio—approximately unity—at all stimulating wave lengths. (2) In addition, a second requirement must be met. For the case of the deuteranope (or, for that matter, the protanope) the sum of the R' and G' ordinates at any wave length must add

to give deuteranopic (or protanopic) luminosity, as specified, for example, by the data of our groups of dichromats.

These relations are shown in Figure 3. The full curve shows color mixture (minus the contribution of a blue fundamental) in the case of a deuteranope, comparable to one of our subjects, who shows no luminosity loss. (The curve is also nearly the luminosity curve for normal subjects.) The right dashed curve represents average deuteranopic luminosity with appropriate loss. In the latter curve, the R' and G' ordinates are taken to be in a ratio of unity, as dictated by color-mixture data. The area under the curve is assumed to be proportional to the number of R' and G' cones (both being identical in all details except their central connections). Each type contributes half the total number. Similar considerations could apply to the logical extension of Fick-Leber ideas to protanopia, as represented in the left dashed-line curve.

The results of the present experiments concerning luminosity losses in dichromats may be at least qualitatively in line with results described by Rushton in another paper of this symposium. For the moment we may regard the main merit of the present accounts as stating how trichromatic theory may account for luminosity losses and the colors seen by dichromats.[10]

* This work was supported by a contract between the Office of Naval Research and Columbia University and by a grant-in-aid from the Higgins Fund of Columbia University. Reproduction in whole or in part is permitted for any purpose of the United States government. For a fuller discussion of this research see C. H. Graham and Yun Hsia, *Science*, **127,** 675–682, 1958.

[1] T. Young, "On the Theory of Light and Colours, in *Lectures in Natural Philosophy* (London: Printed for Joseph Johnson, St. Paul's Church Yard, by William Savage, 1807), **2,** 613–632.

[2] D. B. Judd, *J. Research Nat. Bur. Standards*, **41,** 247–271, 1948.

[3] W. D. Wright, *Researches on Normal and Defective Colour Vision* (St. Louis: C. V. Mosby Co., 1947).

[4] F. H. G. Pitt, "Characteristics of Dichromatic Vision, with an Appendix on Anomalous Trichromatic Vision," *Great Britain Med. Research Council, Special Rep. Series*, No. 200 (1935).

[5] Y. Hsia and C. H. Graham, these PROCEEDINGS, **43,** 1011–1019, 1957.

[6] S. Hecht and Y. Hsia, *J. Gen. Physiol.*, **31,** 141–152, 1947.

[7] C. H. Graham and Y. Hsia, these PROCEEDINGS, **44,** 46–49, 1958.

[8] A. Fick, "Die Lehre von der Lichtempfindung," in *Handbuch der Physiologie*, ed. L. Hermann (Leipzig: Vogel, 1879), **3,** Part I, 139–234.

[9] T. Leber, *Arch. Ophth.*, **15,** 26–107, 1869.

[10] We wish to express our indebtedness to our colleagues Dr. Harry G. Sperling, Mrs. Anne H. Coulson, Dr. Eda Berger, and Mrs. Shakuntala Balaraman for many contributions of time, data, and content to experiments of the program here discussed.

COLOR VISION: THE APPROACH THROUGH INCREMENT-THRESHOLD SENSITIVITY

By W. S. Stiles

NATIONAL PHYSICAL LABORATORY, TEDDINGTON, MIDDLESEX, ENGLAND

I am most sensible of the honor of addressing this meeting and of the generosity of the National Academy of Sciences in bringing me here as their guest. The invitation to contribute to the present symposium came just as I had finished a rather elaborate study of the color-matching properties of the average observer. However, that work I believe to be of more colorimetric interest, and I have elected, therefore, to deal with a quite different approach to color vision by a method which does not rest on actual color-matching—the method of increment thresholds.

Consider, to take first the simplest case, a subject whose eyes are fully adapted to a completely dark field. The increment threshold is then just the smallest perceptible quantity of a certain test light which is applied at a particular part of the visual field. For convenience, I use the expression "smallest perceptible quantity," although, of course, thresholds are definable only statistically, for example as critical values of the stimulus variable at which the chance of a perception equals 50 per cent. In the general case, the increment threshold is still the smallest perceptible quantity of a certain test light, but now this is added to another stimulus—the conditioning stimulus—which consists of any distribution of brightness and color to which we may choose to expose the eye. The conditioning stimulus need not be an unvarying one: the increment threshold is, in fact, often measured while the eye is adapting itself to a change from one constant stimulus to another constant stimulus, the whole spatial and temporal pattern of stimulation being considered to form the conditioning stimulus. The increment threshold (or, better, its reciprocal, the threshold sensitivity) is a measure of visual sensitivity which can be varied in different ways to bring out particular properties of the visual response system. We may, for example, vary the angular size and exposure time of the test stimulus, to determine how well the visual system integrates over area and over time. Most important, we may change the spectral composition of the test stimulus, in order to determine the spectral sensitivities that are active in the response system. So far, only changes in the parameters of the test stimulus have been mentioned. In addition, any number of different conditioning stimuli may be used. However, there is evidence that the modifications of the increment threshold by more complex patterns of brightness and color are similar to those obtained with a very simple pattern, namely, a uniform stimulus which extends over and around the area of the field where the test stimulus is applied. We should regard the action of a more stimulus, including stimuli applied at some distance from the test area, as *equivalent* to that of a suitably chosen uniform stimulus of the kind mentioned if, for all possible test stimuli (of various sizes, colors, exposure times, etc.), the observed increment threshold was the same for the two conditions. The equivalence is probably not generally complete, but it has been found to hold approximately in some cases, and we are justified in concentrating, in the first instance, on the simple stimulus pattern. This has the advantage that much

bigger effects on the increment threshold can be obtained with "covering fields" than with conditioning stimuli confined to the surrounding retina.

For probing the color mechanisms of the eye, the method that has been extensively used is to apply, in flashes, a small monochromatic test stimulus of wave length λ to a uniform monochromatic conditioning field of wave length μ and intensity M_μ extending over the test area (the two-color threshold technique). Most work has been done for the case when the eye is fully adapted to the conditioning field, although several recent investigations (particularly by Boynton[1]) as well as some earlier ones have considered non-equilibrium situations. If all the other possible experimental variables (size and exposure time of test stimulus, angles of incidence on the retina of test, and conditioning stimuli, etc.) are kept constant, the increment threshold N_λ, i.e., the smallest perceptible intensity of the test stimulus, will be a function $f(\lambda, \mu, M_\mu)$ of the three variables λ, μ, M_μ. The function f is certainly not simple, and its determination for the whole range of possible values of the three variables is a considerable undertaking. Some progress has been made, however, which suggests that the function may be put in a form involving separate terms identifiable, on certain reasonable assumptions, with the contributions to the increment threshold of component visual mechanisms having different spectral sensitivities and some other differences in their properties. The possibility of "separating" the function f in this way does not, of course, prove that there exist in the eye separate structures and processes corresponding to the different terms. The psychophysical analysis has to be related to objective observations on the visual pigments, on the types of end-organ and the ramifications of the nerve pathways through the higher neurons, on the electrophysiological response, etc., before the "component mechanisms" can begin to acquire a real existence. Despite the wealth of information on the visual system generally that is provided by these objective techniques, there has until recently been little which could be applied directly to the psychophysical data of color vision. The important new development is the objective detection and measurement of bleachable visual pigments, other than rhodopsin, in the living human eye (Rushton).[2] Some repercussions of this work on the increment-threshold analysis will be mentioned later.

The guiding principle in interpreting increment thresholds is that in the retinal area on which the test stimulus is imaged there are several component mechanisms, any one of which may be the means of triggering off the response "seen" when the test flash is applied. Each component mechanism is conceived of as an association of a selected proportion of the end-organs (rods and cones) in the area, the association being effected somewhere in the neural system through which the nervous activity, initiated by light absorption in the end-organs, is transmitted to the brain and finally to the effector nerves of the external response. It is not assumed that the end-organs of a component mechanism are necessarily all of the same kind or contain just one photopigment, nor that end-organs belong exclusively to one component mechanism, but the mechanisms are assumed to have sufficient independence for the idea of a particular component mechanism acting alone to have a meaning. It is assumed, in fact, that the response of all but one component mechanism could be blocked (ideally; it may not be practicable to do

it), leaving the remaining mechanism to respond in its normal way. For a given test stimulus and given conditioning stimulus, the observed increment threshold in the retinal area is assumed to depend only on the respective increment thresholds of the component mechanisms that would be observed under the same conditions if they were acting alone. In particular, it is assumed that the observed increment threshold will not exceed any of the component increment thresholds and will generally approximate to the smallest of these. Finally, each component mechanism is assumed to have a characteristic relative spectral sensitivity function, so that the effect on the mechanism of an intensity I_1 (in energy units) of one wave length λ_1, forming the whole or part of the test or conditioning stimulus, is just the same as that of an intensity I_2 of any other wave length λ_2 that may be substituted for it, provided that I_1 and I_2 are in the inverse ratio of the relative spectral sensitivity values at these wave lengths.

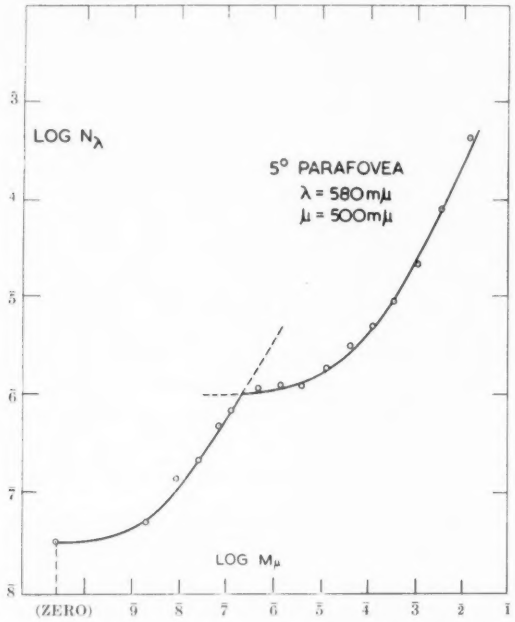

FIG. 1.—Increment threshold, N_λ, versus field intensity, M_μ, for 5° extrafoveal vision. Test stimulus: 1°, 0.06 sec. Subject: W. S. S. ($N\lambda$ and $M\mu$ expressed in erg sec^{-1} [deg. of arc]$^{-2}$.)

The analysis of increment-threshold measurements on the above lines may be illustrated by examining briefly the case of extrafoveal vision, for which there is the ample evidence of the duplicity theory of vision for the operation, in some sense, of distinct rod and cone mechanisms. Figure 1 gives an example of a so-called threshold-versus-intensity (t.v.i.) curve, in which the logarithm of the increment threshold is plotted against the logarithm of the uniform conditioning stimulus for a particular pair of test and conditioning wave lengths, $\lambda = 580$ mμ, $\mu = 500$ mμ. In this case the test stimulus (1°, 0.06 sec.) was imaged 5° from the

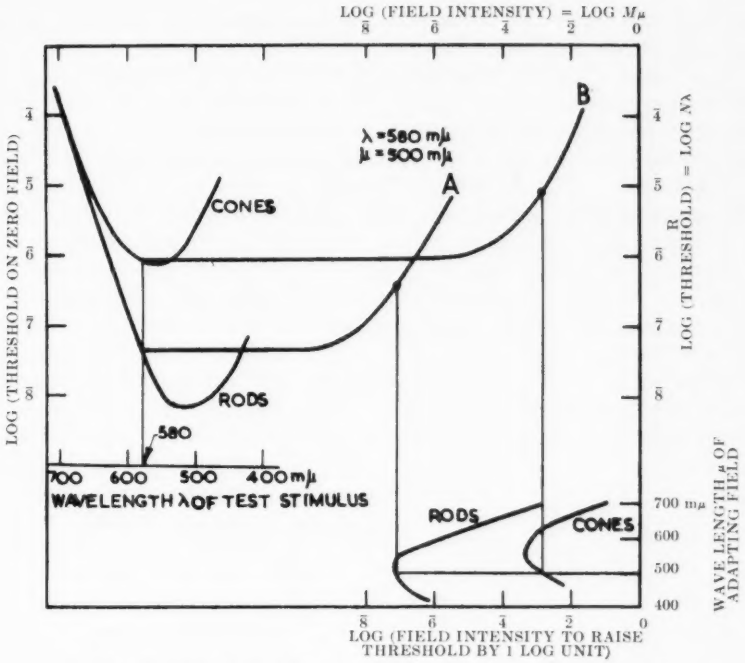

FIG. 2.—Diagram illustrating how t.v.i. curves like that of Fig. 1 result from the t.v.i. curves of two component mechanisms, *A* (rod) and *B* (cone), whose positions with respect to the axes of log $N\lambda$ and log M_μ for any test and field wave lengths λ and μ are determined by the respective spectral sensitivity curves of the mechanisms shown in the auxiliary diagrams.

fovea. Figure 2 shows diagrammatically the tentative interpretation of the curve of Figure 1 in terms of two component mechanisms of the kind contemplated above.[3] Curves *A* and *B* (Fig. 2) represent, respectively, the t.v.i. curves assumed for the *A* (rod) and *B* (cone) component mechanisms when acting alone. At each field intensity the observed increment threshold is put equal to the lower of the values appropriate to the *A* and *B* curves. The crucial question now is What happens if λ and μ are changed? According to the assumption about the spectral sensitivity curve of a component mechanism, a change in λ, keeping μ constant, should merely displace the curves *A* and *B*—without change of shape—parallel to the ordinate axis, by amounts determined by their respective relative spectral sensitivities. A change in μ, keeping λ constant, should, on the other hand, displace the curves parallel to the axis of abscissae by amounts also fixed by the respective spectral sensitivity curves. The spectral sensitivity curves are also shown in the diagram, and it is easy to visualize the effects of changing λ and μ. The experimental t.v.i. curves obtained for various combinations of test and field wave lengths show that the branch of the curve attributed to the *A* mechanism conforms satisfactorily to these *displacement rules*, the displacements parallel to the two axes being controlled by the same spectral sensitivity curve. For the *B* branch, how-

ever, the displacement rules fail to hold, and it is impossible to represent the increment threshold for extrafoveal vision as arising from just two component mechanisms in the way proposed. The next step would be to see whether the B branch could be represented as the resultant of two or more branches, but this is difficult because the form of the B branch for many of the interesting (λ, μ) combinations is not accessible to observation: it is masked by the A branch, which has the lower threshold.

We now examine the various component mechanisms which we are led to postulate for the fovea, the area of the retina where color discrimination is most highly developed and where the absence of rod end-organs eliminates (probably) any participation of a "rod" component mechanism. (This last simplification is not quite a certain a priori, as the adapting field extends into the retinal area containing rods, but it seems to be so in fact.) Consider, first, the t.v.i. curves obtained with test stimuli of short wave length ($\lambda < 510$ mμ) and with conditioning stimuli of medium or long wave length ($\mu > 520$ mμ). These show a division into two branches ($\mu <$ about 560 mμ) or of three branches ($\mu > 560$ mμ), an example of the latter being given in Figure 3. The three branches are provisionally ascribed to three component mechanisms, denoted by the neutral symbols π_4, π_1, π_3 marked in the figure. The changes in the t.v.i. curve when λ and μ are altered establish that the branch π_1 obeys the displacement rules. But if λ becomes too large or μ too small, the π_1 branch is masked by the π_4 branch; thus the displacements observable by changing λ and μ correspond to non-overlapping spectral ranges of the spectral sensitivity curve of the mechanism, and it is not possible immediately to check that a single spectral sensitivity curve determines displacement of both kinds. The difficulty can be largely overcome by the "auxiliary field method." In this we use as conditioning stimulus a mixture of a fixed auxiliary field of long wave length sufficiently intense to bring the increment threshold to the π_1 branch of the curve, and a main field of wave length μ, which may be as short as desired. The comparative effects of main fields of different wave length in moving the increment threshold higher up the π_1 branch can then be determined.[4] It is found that in the overlapping region ($\lambda < 500$ mμ) the λ and μ displacements correspond to a common relative spectral sensitivity curve and that this curve has a maximum at approximately 440 mμ (energy expressed in quantum units). The evidence on the π_3 branch is more limited because of the difficulty in obtaining the very high field intensities needed to reach it. The results again indicate a branch obeying approximately the displacement rules and with consistent λ and μ displacements in the overlap region

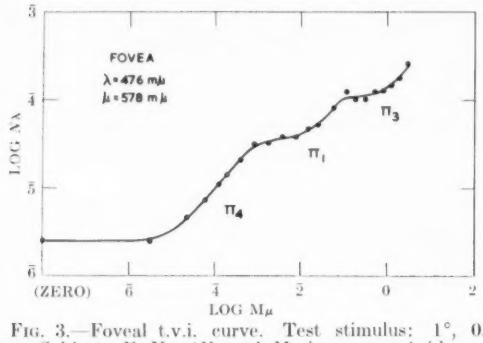

FIG. 3.—Foveal t.v.i. curve. Test stimulus: 1°, 0.2 sec. Subject: E. K. ($N\lambda$ and $M\mu$ in erg sec^{-1} [deg. of arc]$^{-2}$.)

($\lambda < 500$ mμ). As can be seen in Figure 4, the spectral sensitivities of the π_1 and π_3 component mechanisms (derived from μ-variation displacements) agree closely in shape in the region of their maxima ($\lambda < 500$ mμ) and differ only at long wave lengths. (The curves of Figure 4 and later figures refer to the mean results of the same four subjects.) But for test stimuli in the short-wave region the two mechanisms still differ in their absolute thresholds, i.e., the values to which their increment thresholds tend as the conditioning stimulus is reduced to zero: the absolute threshold for π_3 is about four times that for π_1. It was unexpected to find two component mechanisms with maximal and closely similar spectral sensitivities in the blue. But there is also some evidence of a third component mechanism (π_2) with maximal sensitivity in the same spectral region. This is indicated by a breakdown in the displacement rules for the branch π_4 when, for conditioning fields of long wave length (e.g., 600 mμ), the wave length of the test stimulus is reduced below about 460 mμ. The discrepancy would be consistent with a π_2 mechanism with the spectral sensitivity curve (by the μ-variation method) in the range 510–700 mμ, as shown in Figure 4. For $\mu <$ 510 mμ, the curve must lie above the points of the arrows shown in the figure. Unfortunately, the auxiliary field device is not applicable in this case, and the form of the curve has not been determined in the neighborhood of the maximum, although the latter can hardly lie outside the range 440–480 mμ.

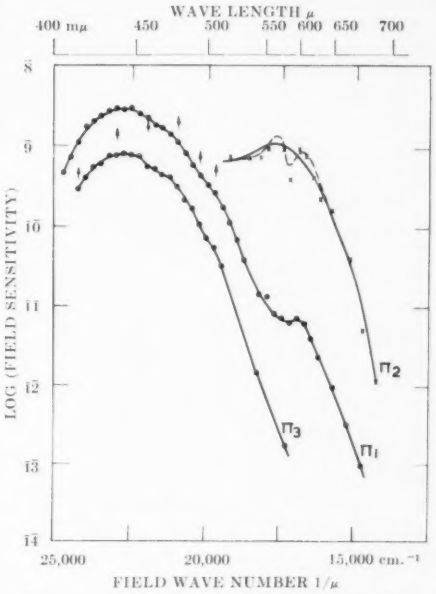

FIG. 4.—Field spectral sensitivity of foveal cone mechanisms π_1, π_2, and π_3. Field sensitivity is defined as the reciprocal of the field intensity in quanta sec^{-1} (deg. of arc)$^{-2}$ required to raise the increment threshold of the mechanism to ten times its zero-field value. For $\lambda < 510$ mμ, the arrow points indicate values below which the undetermined part of the π_2 curve cannot fall. Means for four subjects, π_1 and π_3; for three subjects, π_2.

As the wave length λ of the test stimulus is increased above about 450 mμ, the branches of the t.v.i. curve associated with mechanisms having maximal sensitivity in the blue move upward with respect to the π_4 branch, and for λ greater than 520 mμ they move completely above it and are inaccessible to measurement, whatever the wave length of the conditioning stimulus. For these test stimuli of longer wave length, the π_4 branch does not obey the displacement rules, although the well-marked transitions obtained at shorter wave lengths are not in evidence. It was originally thought that the results would be explained by admitting that there were, in fact, two branches—a π_4 and a π_5 branch—with associated spectral sensitivities not very widely separated along the spectrum, so that sharp transitions from one

branch to the other would not be expected. Further measurements indicated, however, that, while this explanation is fairly satisfactory if attention is confined to conditioning stimuli below about 150 trolands, it leads to discrepancies with observations made at higher intensities. The spectral sensitivities of π_4 and π_5 derived by the μ-variation method from measurements in the lower-intensity range are shown in Figure 5. These sensitivities, like those of Figure 4, represent at each wave length the reciprocal of the field intensity (in quantum units) required to raise the increment threshold of the particular mechanism to ten times its zero-field value. If we also know the shape of the branch of the t.v.i. curve appropriate to the mechanism and the absolute threshold of the mechanism for some test wave length, we can derive the curve showing the variation of the increment threshold or its reciprocal, the threshold sensitivity, with test wave length for any intensity and wave length of the field. This calculation has been made in Figure 6 for the five component mechanisms π_1–π_5 for the case of a field of wave length 667 mμ and intensity 29 trolands. The observed threshold sensitivity at each test wave length should equal approximately the highest of the component threshold sensitivities, and it is apparent that only π_4 and π_5 are concerned in this instance. The circle points represent the directly observed threshold sensitivities and are in satisfactory agreement with the prediction. The comparison of the predicted and observed curves when the field intensity is raised by a factor of nearly 100 is shown in Figure 7. The discrepancy is not only in the absolute values (which might arise from uncertainty in the shape of the π_4 branch) but also in the shape of the observed curve in the wave-length range where it should be at least parallel to π_4. The indication is that, with a high-intensity red field, the π_4 mechanism is relatively more sensitive on the short-wave side of about 580 mμ than the calculation treating π_4 as a single component mechanism with its low-intensity properties would predict. This deviation is independently confirmed by t.v.i. curves taken to high intensities. An analogous, although smaller, effect for π_5 is shown by observations with low- and high-intensity green conditioning fields (Figs. 8 and 9) and is again confirmed independently by t.v.i. curves. The π_5 mechanism under high green stimulation is relatively more sensitive on the long-wave side of about 580 mμ than

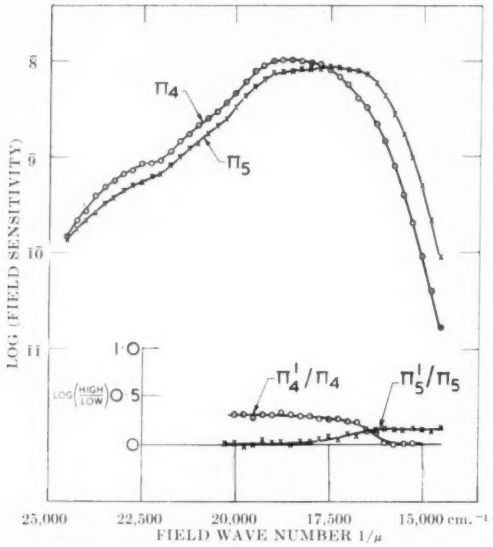

FIG. 5.—Field spectral sensitivities as in Fig. 4 for cone mechanisms π_4 and π_5. The inset diagram shows the log (ratio of the sensitivities: high-intensity level/low-intensity level) for the π_4 and π_5 mechanisms. Means for four subjects.

the calculation treating π_5 as a single component mechanism with its low-intensity properties would predict. Insufficient is yet known of these deviations from single component properties of π_4 and π_5 to say whether they could be explained by assuming additional component mechanisms. For the present, it is advisable merely to recognize modified high-intensity conditions of these mechanisms, denoted by π_4' and π_5', respectively. The differences in spectral sensitivity in the low- and high-intensity conditions of π_4 or π_5 are not large compared with the differences between π_4 and π_5. A first determination of the ratio of the high- to the low-intensity spectral sensitivity is shown in the inset diagram of Figure 5. A tentative deduction

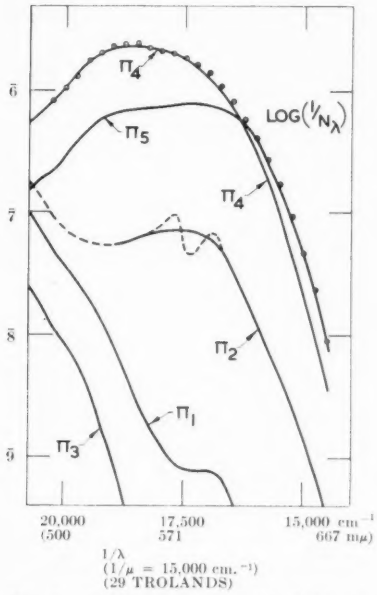

Fig. 6.—*Points:* Increment-threshold sensitivity by $1/N\lambda$ versus test stimulus wave number 15,000 cm^{-1} (wave length 667 mμ) and intensity 29 trolands. Test stimulus: 1°, 0.2 sec. Means of four subjects. *Curves:* predicted threshold sensitivity versus wave-number curves at this conditioning stimulus for the component mechanisms π_1–π_5.

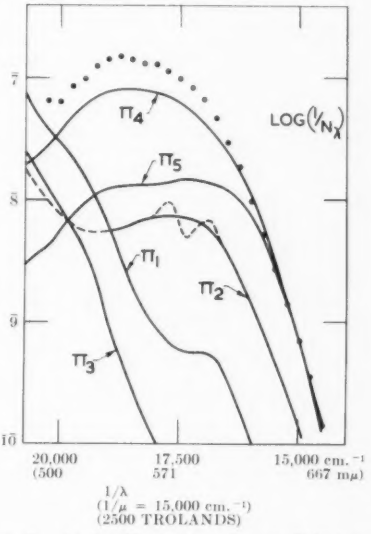

FIG. 7.—As for Fig. 6, but for conditioning field of intensity 2,500 trolands.

from other observations is that the transitions $\pi_4 \rightarrow \pi_4'$ and $\pi_5 \rightarrow \pi_5'$ occur in a range of conditioning intensities of the order of 1 log unit centered on an intensity of about 300 trolands.

The scheme of component visual mechanisms indicated by increment-threshold analysis is summarized in Table 1. The Fechner fraction quoted in the final column is defined as the ratio of the increment threshold to the field intensity when the conditioning field has the same color as the test stimulus ($\lambda = \mu$) and the increment threshold is raised by it to ten times the zero-field value. There are interesting differences between the Fechner fractions of different component mecha-

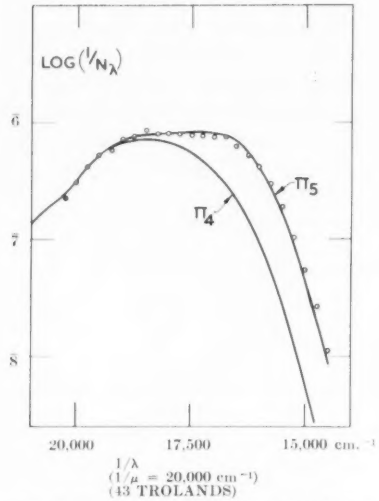

FIG. 8.—As for Fig. 6 but for conditioning field of wave number 20,000 cm^{-1} (wave length 500 mμ) and intensity 43 trolands.

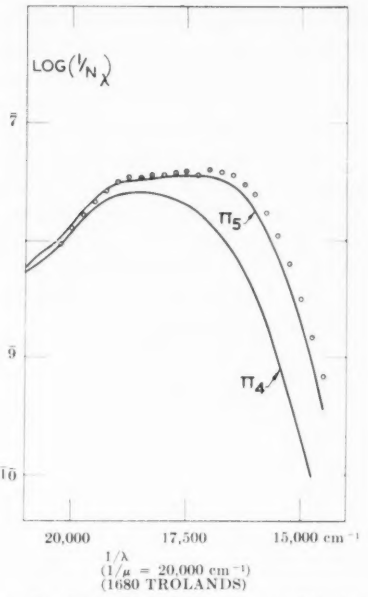

FIG. 9.—As for Fig. 8 but for conditioning field of intensity 1,680 trolands.

nisms: the "blue" cone mechanisms have values four to five times those for the "green" as "red" cone mechanisms. In applying the present ideas of component mechanisms to the explanation of other psychophysical measurements on color discrimination and brightness matching, this difference plays an important part: it corresponds to the small contribution to luminance of short-wave stimuli, despite their strong coloring value.

TABLE 1

Mechanism	Symbol	Remarks	Wave Length of Maximal Sensitivity (mμ)	Fechner Fraction (1°, 0.2 Sec. Test Stimulus)
Rod	π_0	Absent at the fovea	503	(30)
"Blue" cone	π_1	Ratio of blue to yellow sensitivity (440/590) approx. 2.6 log units	440	8.7
	π_2	Ratio of blue to yellow sensitivity (440/590) order of 1 log unit	? (Between 440 and 480 mμ)	? (Probably as for π_1 and π_3)
	π_3	Ratio of blue to yellow sensitivity (440/590) approx. 4.0 log units	440	8.7
"Green" cone	π_4		540	1.9
	π_4'	Modified high-intensity state	540	...
"Red" cone	π_5		575 (very flat max.)	1.8
	π_5'	Modified high-intensity state	587	...

The discussion so far has been concerned with the breakdown of the three-variable function $f(\lambda, \mu, M_\mu)$, all the parameters of the test stimulus other than color being assumed constant. If one of these is allowed to vary, the effects on the different branches of the t.v.i. curves can be compared, and more information on the properties of the component mechanisms can be obtained. For example, it is known that the response of the cone end-organs of the retina is highly directional, and, by determining the change in increment threshold on altering the angle of incidence on the retina of the rays forming the test stimulus (keeping, of course, the retinal area stimulated the same), differences in the directional properties of different mechanisms can be studied. The differential effects are small, and more measurements are needed, but the early work indicates that π_1 is more directional in its reponse than π_4, that π_4 is more directional than π_5 for stimuli in the green and yellow, but that this position is reversed for red stimuli. Changes in the relative response of different mechanisms as the angle of incidence is varied may be expected to produce changes of color even of monochromatic stimuli. The small changes, observed are, in fact, in general agreement with the relative directional properties of the mechanisms obtained by the threshold observations.

Two outstanding questions are: (*a*) What is the relation of the increment-threshold mechanisms to the three processes of trichromatic color-matching? and (*b*) Do the spectral sensitivities of the component mechanisms represent photosensitivities of visual pigments? We limit ourselves to foveal vision, so that the complication of the rod mechanism does not arise. The two questions are closely related, as the three small-field color-matching functions which sum up the color-matching properties of the foveal retina (conditions of very high intensity excluded) are almost certainly linear combinations of the spectral photosensitivities of three visual pigments, although the coefficients in these combinations are not determinable by straightforward color-matching measurements. Thus, if the spectral sensitivity of a component mechanism could be represented as a linear combination of the color-matching functions, it would also be a linear combination of the pigment photosensitivities. This would indicate that the end-organs of the mechanism contained the pigments represented in the linear combination (i.e., those having non-zero coefficients) and that the effects of light absorption in these different pigments added up linearly. Electrophysiological evidence makes it unlikely that this last state of affairs could come about unless all end-organs of the mechanism contained the pigments in a mixture of the same composition, including the specially simple case of a "mixture" comprising just one pigment. The rigorous application of the argument is hampered by the difficulty of obtaining "hard" data from increment-threshold measurements. However, tests made on the mean spectral sensitivity curves (Figs. 4, 5) indicate that π_4 or π_4' can be expressed fairly well as linear combinations of the small-field color-matching functions. For π_5 the linear representation is less satisfactory, particularly for the high-intensity condition π_5'. For π_3 and for π_1 in the main (short-wave) region of the curve, a linear representation is acceptable, and even with the inclusion of the long-wave lobe on the π_1 sensitivity curve it could not certainly be excluded. The actual shapes of the sensitivity curves also have a bearing here. The spectral absorption and spectral photosensitivity curves that have now been obtained for many visual pigments from animal retinas, by measurements on retinal extracts or complete excised

retinas, differ little in shape from the absorption curve for the pigment of rod vision, rhodopsin; to a first approximation, they correspond to the rhodopsin curve (log absorption factor against wave number) displaced along the wave-number axis so that its maximum occurs at the appropriate position. Before physical absorption and photosensitivity curves can be compared with psychophysical sensitivity curves, the latter must be corrected for the selective light losses in the eye that precede the actual absorption by the photosensitive pigment. The correction is uncertain in the violet but is probably not seriously in error in the blue and green: it is negligible at longer wave lengths. It is found that the corrected π_4, π_4', and π_3 spectral sensitivity curves and the π_1 curve ignoring the long-wave lobe are fairly well represented by displaced rhodopsin absorption curves (Fig. 10: the broken lines are, in fact, the corrected spectral sensitivity curves of the rod mechanism, but the latter has substantially the same shape as the rhodopsin absorption curve in the relevant spectral region). An attempt to represent the complete π_1 curve by the addition of a second rhodopsin-shaped absorption curve whose absorption combines linearly with that of the first is not quite acceptable: the linear summation produces too smooth a transition. But it is for π_5 and π_5' that the representation is least satisfactory: neither a simple rhodopsin-shaped curve nor a linear summation of two such curves (however displaced along the spectrum) will represent the spectral sensitivity of the "red" cone mechanism in low- or high-intensity conditions. It seems certain that the end-organs of π_5 cannot be equipped either with a single rhodopsin-type pigment or with a simple mixture (the same in all end-organs) of two such pigments. It has already been noted that the π_5 and π_5' spectral sensitivities are not adequately represented by linear combinations of the small-field color-matching functions, and the failure to fit rhodopsin-type curves might result from the fact that the increment-threshold spectral sensitivity is not built up linearly from the spectral absorption curves of the pigments contained in the end-organs of the π_5 mechanism. It is most improbable, however, that this

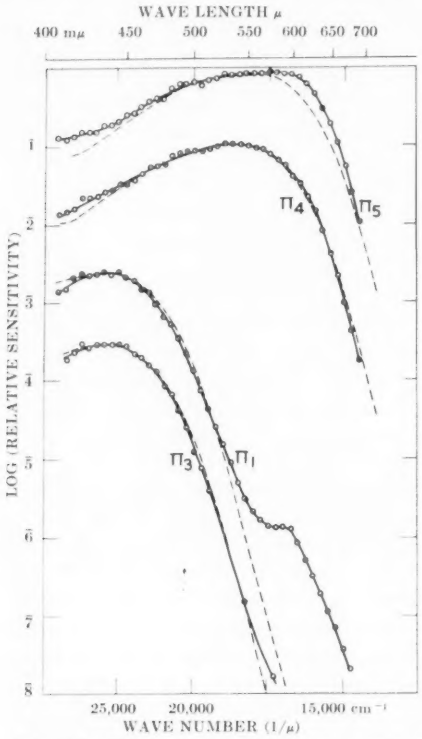

FIG. 10.—Field spectra sensitivity curves of mechanisms π_1, π_3, π_4, π_5 corrected for pre-receptor light losses and compared in each case with the similarly corrected sensitivity curve of rod vision displaced parallel to the two axes to give the best fit.

is the sole cause. Other proposed fundamental (pigment) spectral sensitivities for color vision obtained as strict linear combinations of the color-making functions for which the coefficients have been derived from the properties of the three kinds of dichromatic color defectives and other psychophysical measurements demand a similar non-rhodopsin-type "red" curve, usually including still more striking deviations from the rhodopsin shape at short wave lengths (Fig. 11).

Rushton's[2] objective measurements of bleachable pigments in the human fovea have established the presence in protanopes, deuteranopes, and color normals of a pigment he has named *chlorolabe* with a difference spectrum and action spectrum having a maximum at 540 mμ. As he has pointed out, the spectral sensitivity of the π_4 mechanism is a fair approximation to the spectrum of chlorolabe. The

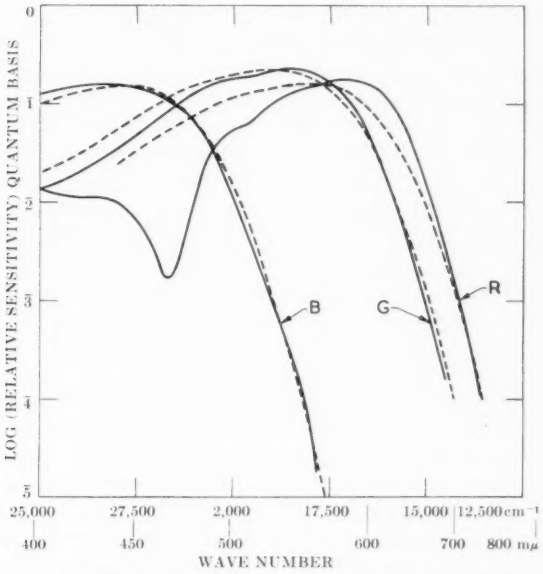

FIG. 11.—Thomson-Wright colour fundamentals corrected for light losses and compared with the displaced rod curve as in Fig. 10.

π_4' spectral sensitivity curve lies still closer to the objective curve. In fact, we have here a convergence, onto a common curve of rhodopsin type, of four spectral sensitivities: the objective spectrum of chlorolabe, the "green" fundamental derivable from the small-field color-matching functions of normal and dichromatic subjects,[5] the increment-threshold sensitivity π_4', and, finally, the sensitivity of a normal eye in a state of "artificial monochromasy" produced by exposure to very intense violet and red stimuli (Brindley[6]). The implication for the "green" cone mechanism is that, in the high-intensity condition, its end-organs contain, or at least are activated by the light absorption of, just one pigment—chlorolabe. Why, then, should the spectral sensitivity π_4, attributed to the mechanism in the low-intensity condition, be different? If exactly the same end-organs were concerned

at all intensity levels, either these would have to contain—probably in relatively small amount—a second pigment more red-sensitive than chlorolabe, or, alternatively, the effective spectral absorption curve of the latter pigment in its situation within the end-organs would have to vary with the intensity of stimulation. A variation of the kind required would occur if at low levels of stimulation the optical density of the pigment in the end-organ was high enough for self-screening to occur and if the reduction of density by bleaching at high levels of stimulation substantially reduced the self-screening. "Self-screening" denotes here the selective filtering of the light reaching the deeper pigment layers in the end-organ, in its passage through the upper pigment layers. Brindley has considered the possibility that the breakdown of metameric color-matches after the eye has been exposed to high intensities may be explained by the reduction of self-screening present at low and moderate intensities.[6] The difference between the π_4 and π_4' spectral sensitivities would correspond—very approximately—to a change in self-screening if in the low-intensity condition the chlorolabe in the end-organs had an optical density at the maximum wave length 540 mμ of about 1.0 and if this was reduced to about 0.2 in the high-intensity condition. However, the existence of self-screening in the end-organs of the green cone mechanism is contrary to all the evidence: Brindley's results show that the breakdown in metameric color-matches in the green-yellow-red range is associated with an anomaly not of the "green" process but of the "red" process only; the transition from π_4 to π_4' occurs in an intensity range lower than that required to produce appreciable breakdown in color-matches; the requisite density of the chlorolabe under zero-field conditions and the necessary degree of bleaching in passing from the π_4 to the π_4' intensity level are both improbably high, according to Rushton's objective data on chlorolabe. As a more red-sensitive pigment, *erythrolabe*, has been shown objectively to be present in the retina (Rushton),[2] the suggestion that the end-organs of the "green" cone mechanism might contain a small proportion of it mixed with chlorolabe is, at first sight, promising. The difficulty here is to understand why the red-sensitive pigment should be bleached or rendered ineffective by an increase in intensity corresponding to the transition π_4 to π_4'. The objective data on erythrolabe would poind to a comparatively small reduction—not more than 10 per cent—in its density by bleaching in this intensity range. We are led, therefore, to the other main possibility that, at low intensities, additional end-organs, not containing only chlorolabe, have contributed to the spectral sensitivity derived for the "green" cone mechanism. Many ways can be imagined in which this might come about, but it would be pointless to enlarge on them, as there is at present little evidence to distinguish different suggestions. One general observation may be made: whatever the way in which the non-chlorolabe end-organs are associated with the chlorolabe end-organs in determining π_4, the association is broken as a result of increasing the level of stimulation.

It is possible that in three blue-sensitive mechanisms—π_2, π_1, π_3—we have something like the same principle in operation. If we accept π_3 as the "blue" cone mechanism with end-organs containing just one blue-sensitive pigment of rhodopsin type (cyanolabe? not yet detected objectively), the spectral sensitivities of the other mechanisms would correspond to associations of cyanolabe with non-cyanolabe end-organs, these associations being broken at higher levels of stimulation by long-wave light. On this view, only one of the three mechanisms would be in

being at any moment, and this represents formally a radical change in our working hypothesis. But, while it can be shown from the steady-state increment-threshold measurements that the mechanisms π_1 and π_4 are present together, the relationship of the spectral sensitivity curves of π_1, π_2, and π_3 prevents a similar proof that any two of these are simultaneously present.

Comments generally similar to those made on π_4 and π_4' apply to the change from π_5 to π_5', but there are some differences. There is an additional objection to the use of self-screening as an explanation of the change; bleaching of the self-screening pigment would lead to a change in the spectral sensitivity curve even qualitatively different from that observed. This is not necessarily in conflict with Brindley's conclusion that, at higher-intensity levels, a change in self-screening of the end-sensitive pigment accounts for the breakdown of metameric color-matches, although the objective data on the erythrolabe in the retina do not support this theory. There is no convergence of the different spectral sensitivities with maxima at long wave length on a common rhodopsin-type curve. The π_5 curve does not agree with the objective spectrum of erythrolabe: Rushton has shown that π_5 can be fitted by the absorption curve of a mixture of erythrolabe and chlorolabe in the proportion 1:0.6.[2] For π_5', the corresponding proportion would be about 1:0.3, so that π_5' is nearer the objective curve of erythrolabe. The fact that both erythrolabe and chlorolabe have been shown to be present in the deuteranopic eye favors the Fick hypothesis of red-green fusion for the explanation of deuteranopia. The character of dichromatic color vision as a reduction form of normal color vision means that the spectral sensitivity of the red-green fusion in the deuteranope must be a linear combination of the normal's color-matching functions and hence almost certainly of the spectral absorption (strictly photosensitivity) functions of the normal's pigments. Deuteranopia would then be explained by the presence in all the end-organs of the fused red-green system of a mixture in an invariable ratio of the chlorolabe and erythrolabe pigments; any other explanation would encounter linearity difficulties. If, as is sometimes assumed, the normal also has the fusion process of the deuteranope but, in addition, the simple "green" process (and, of course, the "blue" process also possessed by the deuteranope), then the appearance in the normal eye of a π_5–π_5' mechanism supplied by end-organs containing a mixture of chlorolabe and erythrolabe would fit in very well. But there are difficulties: on the one hand, the pigment mixture appears unexpectedly stable when very high-intensity colored stimuli are used in increment-threshold and artificial monochromasy measurements, and, on the other hand, the breakdown of color-matches at high-intensity levels could not correspond merely to an alteration of the composition of a pigment mixture in the end-organs of the "red" cone mechanism.

This review has been devoted mainly to the spectral sensitivities of the color-vision mechanisms of the fovea arrived at by increment-threshold measurements made under steady-state conditions. The spectral sensitivities are perhaps of special interest at this time in the light of the new objective data on cone pigments. The summation properties of different color mechanisms, the way the mechanisms are modified in moving out from the fovea, and their behavior when the conditioning stimulus is suddenly changed have all been studied by the increment-threshold method to a limited extent. There is an interesting field here for further work.

[1] R. M. Boynton, *J. Opt. Soc. Amer.*, **46**, 172, 1956.

[2] W. H. A. Rushton, *Visual Problems of Colour* (N.P.L. Symposium, No. 8) (in Press).
[3] W. S. Stiles, *Documenta Ophthal.*, **3**, 138, 1949.
[4] W. S. Stiles, *Coloquio sobre problemas opticos de la vision* (Madrid: Union Internationale de Physique Pure et Appliquée, 1953), **1**, 65.
[5] L. C. Thomson, and W. D. Wright, *J. Opt. Soc. Amer.*, **43**, 890, 1953.
[6] G. S. Brindley, *J. Physiol.*, **122**, 332, 1953.

VISUAL PIGMENTS IN THE INTACT HUMAN EYE

BY W. A. H. RUSHTON

CAMBRIDGE UNIVERSITY, CAMBRIDGE, ENGLAND

It is generally agreed that the first step in the visual process is the absorption of light by one or more photosensitive visual pigments contained in the rods and cones of the retina. The pigment of the *rods* has long been recognized and studied under the name *rhodopsin*, but rod vision is the colorless vision of twilight, and thus, to understand the photochemical basis of color vision, we need to measure the visual pigments in the *cones*. No extract of mammalian eyes has ever yielded a visual pigment other than rhodopsin; this is not surprising in view of the observations of Wald, Brown, and Smith[1] that the chick's eye, where cones relative to rods are a thousand times as abundant as in man, still contains twice as much extractable rhodopsin as iodopsin (the cone pigment). If in man we are to measure the pigments in cones, we must take advantage of two exquisite features of retinal organization. First, there is a tiny area of the retina—the *fovea centralis*—0.6 mm. in diameter, where there are no rods and where the cones are close-packed. If measurements can be confined to this region, contamination by the enormous preponderance of rhodopsin will be absolutely excluded.

The difficulties in dealing with so small a sample are somewhat offset by the second feature of cone structure, for cones are beautifully organized to catch quanta and secure photolysis. By measuring the pigments *in situ* within the cone, we take advantage of this organization and not only detect physically the change in pigment density during bleaching and regeneration but measure this change to an accuracy of about 5 per cent by a procedure which takes about 7 seconds to perform.

In principle, the method is to analyze the light reflected from the eye as seen in a cat's eye on the road by night or in man through an ophthalmoscope. This reflected light has passed twice through the retina and so must have suffered absorption by the retinal pigments traversed. If the incident light falls only upon the tiny *fovea centralis*, the pigments will be cone pigments, and if these are bleached to transparency by a strong dazzling flash, the reflected light will return brighter. Thus measurement of the increase in intensity of lights of various wave lengths will give the *difference spectrum* of the cone pigments involved.

Preliminary reports of equipment and results have already appeared,[2] and some quantitative results upon the kinetics of cone pigments have been published.[3] It will therefore be unnecessary to repeat here the *results* presented at the National Academy, though they show some improvement upon the earlier experiments. But the main conclusions may be summarized.

1. *Measurements upon the protanope.*—This is the type of red-green color blind who cannot see the far red spectrum at all.

a) He exhibits a change in foveal reflectivity upon "bleaching" with a bright light and subsequent "regeneration" in the dark, such as would be expected if what we were measuring was the density of a visual pigment.

b) The amount of bleaching produced is related to the energy of the bleaching light by the curve expected from photochemical theory.

c) Bleaching by red or green lights caused identical difference spectra; thus the protanope has only one visual pigment (in the red-green range). This would account for his monochromacy in this range if the pigment measured were that upon which vision depends.

d) The pigment measured is a *visual* pigment because red and green lights, which appear to the protanope identical, are found to bleach equally. And the difference spectrum coincides well with the protanope's luminosity curve.

2. *Measurements upon the normal eye.*—

a) A *deep* red bleaching light, which has no effect at all upon protanope pigments, gives, with the normal, a difference spectrum, indicating the presence of a red-sensitive pigment which the protanope lacks.

b) The total bleach in the normal shows the presence of a second photosensitive pigment, and this appears identical with that found in the protanope. Names for these two human cone pigments are "chlorolabe" (= green-taking), with maximum absorption at 540 mμ, and "erythrolabe," with maximum at 590 mμ.

c) In the normal there is somewhat less chlorolabe than erythrolabe, but in the protanope the amount of chlorolabe is about equal to the sum of all the pigments in the normal eye. It therefore seems that the protanope who has the normal number of cones is peculiar just in this—that those cones which in normal men contain erythrolabe have this pigment replaced by chlorolabe in the protanope.

[1] G. Wald, P. K. Brown, and P. H. Smith, *J. Gen. Physiol.*, **38**, 623–681, 1955.
[2] W. A. H. Rushton, *N.P.L. Symposium*, Paper I (London: H.M. Stationary Office, 1958).
[3] W. A. H. Rushton, *Proc. N.Y. Acad. Sci.*, **74**, 291–304, 1958.

---

*COLOR VISION AND THE NATURAL IMAGE. PART I**

BY EDWIN H. LAND

POLAROID CORPORATION, CAMBRIDGE, MASSACHUSETTS

We have come to the conclusion that the classical laws of color mixing conceal great basic laws of color vision. There is a discrepancy between the conclusions that one would reach on the basis of the standard theory of color mixing and the results we obtain in studying total images. Whereas in color-mixing theory the wave-lengths of the stimuli and the energy content at each wave-length are significant in determining the sense of color, our experiments show that in images neither the wave-length of the stimulus nor the energy at each wave-length determines the color. This departure from what we expect on the basis of colorimetry is not a small effect but is complete, and we conclude that the factors in color vision hitherto

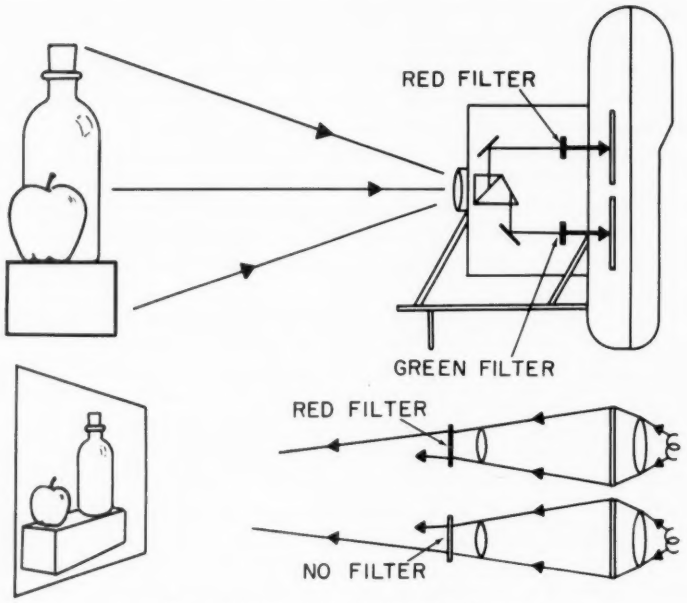

FIG. 1.—Diagram of camera and projector

regarded as determinative are significant only in a certain special case. What has misled us all is the accidental universality of this special case.

It was scientifically proper to base our earlier hypotheses about the functioning of the eye on our knowledge from this special case. Now, however, these recent experiments show that the special case is extraordinarily restrictive: The eye will see color in situations entirely unpredictable on the basis of older hypotheses.[1] Thus it may be necessary to build a new hypothetical structure to explain the

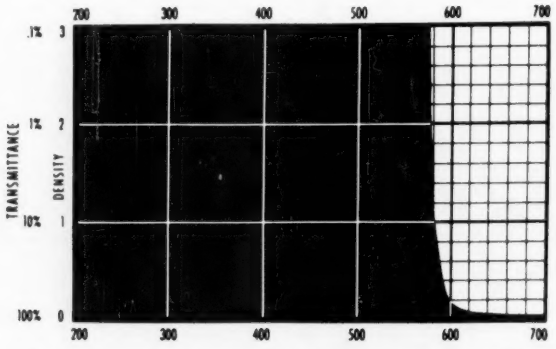

FIG. 2.—Kodak Wratten filter No. 24. Diagram of filter transmittance reprinted, with permission, from the copyrighted Eastman Kodak Company publication, *Kodak Wratten Filters for Scientific and Technical Use*. This applies also to Figs. 8, 9, 10, and 11.

mechanism for seeing color. The purpose of this paper, however, is not to discuss internal visual mechanisms, but rather to begin the description of the laws for a large *general* case—laws describing the relationship of the sense of color in the human eye to the structure of *total images*, laws in which wave-length of stimulus and intensity of stimulus will play entirely new roles.

APPARATUS

For these experiments two black-and-white photographic positive transparencies are made in a split-beam camera, so that they are taken at the same time from the same viewpoint through the same lens (Fig. 1). One is taken through a Wratten No. 24 filter, passing wave-lengths longer than about 585 mμ (Fig. 2), and one through a Wratten No. 58 filter, passing wave-lengths shorter than about 585 mμ. These two photographic records will be referred to hereafter as the "long record" and the "short record," respectively. The picture is exposed so that the gray scale appears to have the same densities in both records.

FIG. 3.—Blonde girl (Experiments 2, 3, 4, and 5). Upper image projected with longer wave-length stimulus; lower image with shorter wave-length stimulus. This applies also to Figures 4, 5, 6, 7, 12, and 13.

The two records are projected in a double-lens projector and superimposed exactly on a screen by means of a registration system. In front of each lens is a pair of polarizers mounted so that the intensity of the light in each beam can be varied separately. In projecting, the long record is projected ordinarily through a red filter (Wratten No. 24), and the short record through a neutral density filter of about 0.3 (when the polarizers are not being used). The half of the projector containing the long record will be referred to hereafter as the "long projector," and the other as the "short projector."[2]

CHROMATIC PHENOMENA

The experiments demonstrated were the following:

Experiment 1.—The double projector is turned on, a filter transmitting red

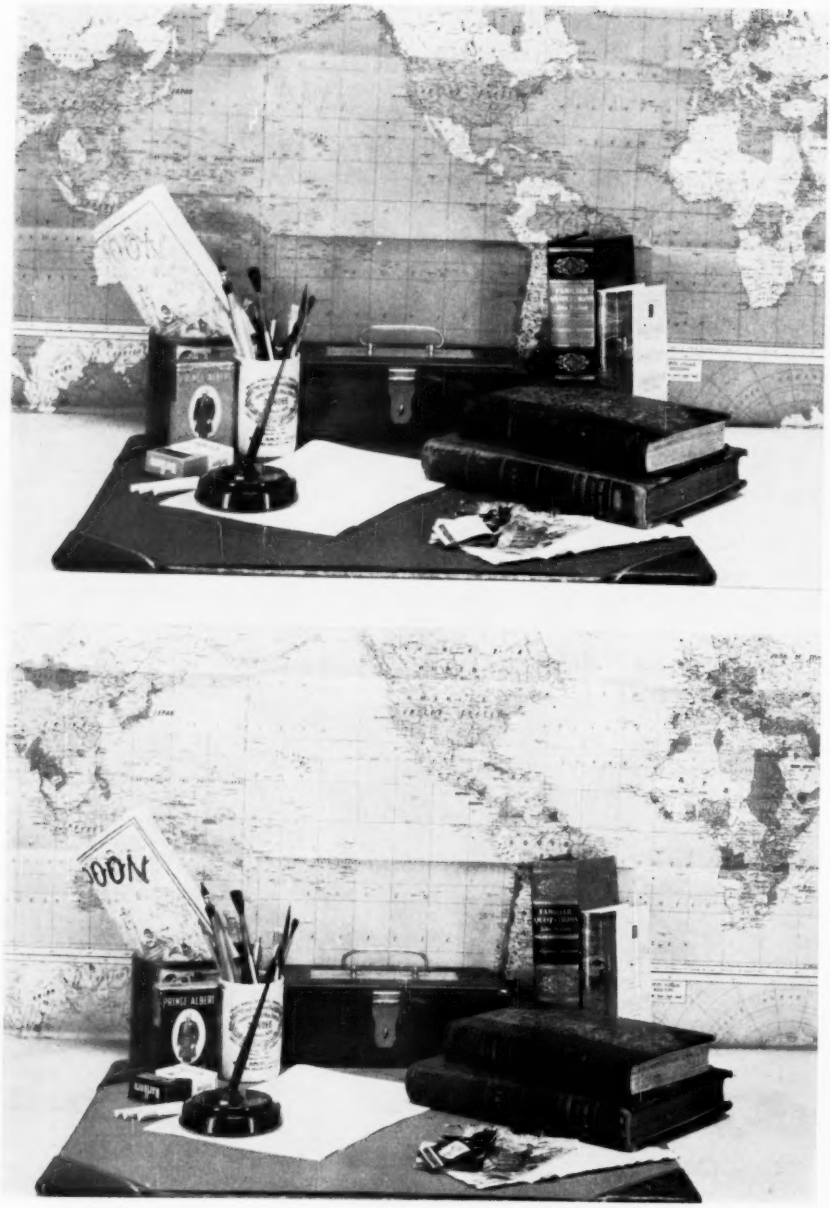

FIG. 4.—Desk and map (Experiments 6 and 20)

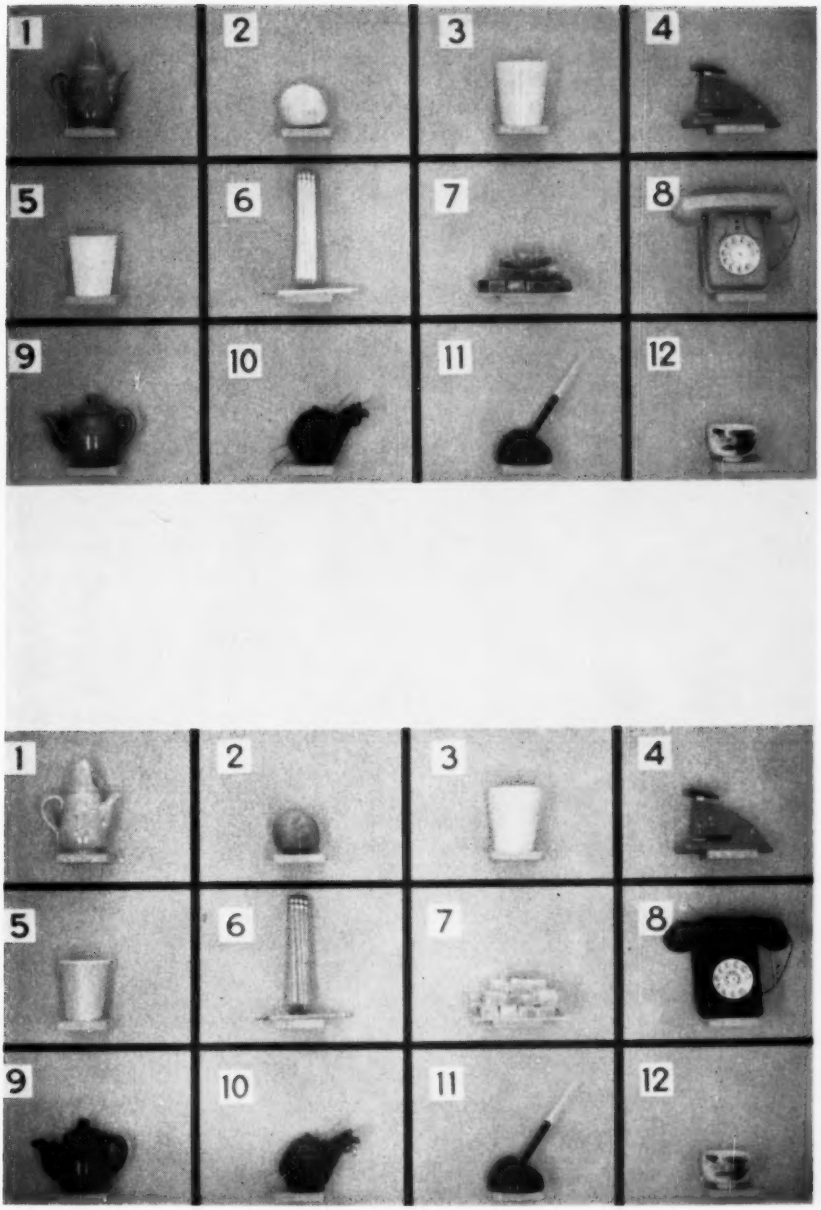

FIG. 5.—Numbered board of standard objects (Experiments 9, 10, and 11)

(Wratten No. 24) is placed over the lens of one projector and the polarizers are rotated to change the observed color of the screen from full white to full red, traversing the various shades of pink in between. Thus the screen is seen in the colors expected on the basis of classical color theory.

Experiment 2.—The red filter is removed, and the two components in Figure 3 are shown separately by occluding first one of the projectors and then the other, so that the audience may see the two black-and-white images produced by the beam-splitting camera. The differences to be expected in any color-separation photographs between the picture taken through the red filter and the picture taken through the green filter are pointed out to the audience.

FIG. 6.—Groceries (Experiment 12)

Experiment 3.—With the short projector occluded, the long record is shown projected through the red filter. Then the lens of the short projector is uncovered so that the black-and-white picture is combined with the red picture. Instantaneously the portrait appears in full color: blonde hair, pale blue eyes, red coat, blue-green collar, and strikingly natural flesh tones. (While it is not the primary purpose of these experiments to achieve verisimilitude, one is impressed by the frequent acceptability of this and many of the other pictures as lovely color photographs.)

Experiment 4.—The room lights are turned on full, flooding the screen with light, and it is shown that there is no visible time lapse between snapping off the lights and the appearance of full color in the picture.

Experiment 5.—The room lights are brought up from darkness to a medium level at which objects in the room can be clearly seen in color, but at which there is not so much light that the picture is washed out entirely. It is shown that the eye sees color in the room and color on the screen at the same time and that, furthermore, the colors on the screen are not changed by having the surrounding room illuminated in white light. This experiment gives the first premonition that multiple color universes can coexist side by side, or one within another.

Experiment 6.—In this picture (Fig. 4) the audience is to compare the pink states on the map with the yellow pencil; and the yellow pencil with the white pencil and the white marmalade jar which holds the pencils. It is pointed out that since

classically all the colors on the screen would be pink, it is particularly important to note the difference between the pinks and the yellows and equally important to notice the strong sense of whiteness in the white objects. This is the first demonstration that it is not necessary to use traditional complementary colors to produce white, a fundamentally important fact in the image situation. The green-blue color of the large book entitled *Bartlett's Familiar Quotations* is contrasted with the deep-blue stripe on the air-mail letter, as well as to the blue tax stamp and the pale blue ocean on the map. The gamut of colors in this picture includes also the brown wooden box, the green blotter, the black inkstand, and the red lettering on the magazine. It is pointed out that this slide demonstrates dramatically that the transition from one stimulus (here, either red or white) to two (here, red *and* white) is the tremendous step: the full gamut of color appears. (The addition of a third stimulus does enrich the colors somewhat, but we do not know whether what is missing is due to inadequacy of technique or to the need of a little help from a third stimulus.)

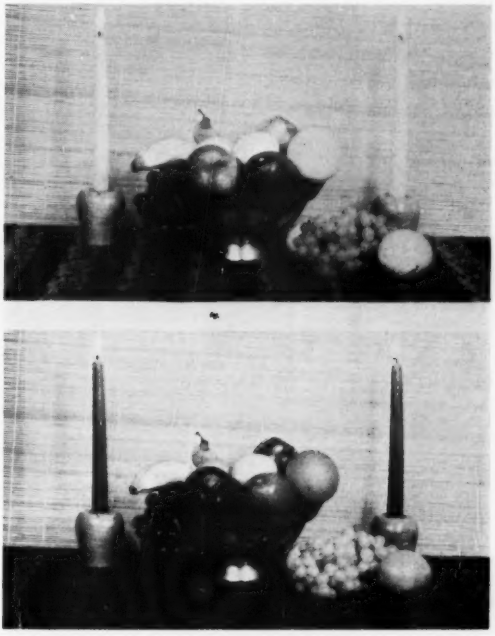

FIG. 7.—Bowl of fruit (Experiments 13, 14, 15, and 16)

Experiment 7.—A colorful picture, notable for the auburn red hair and the green sweater, is used to show that the field that counts in the image situation is not determined by angular subtend. The observer finds the sweater to be green even when he stands so close to the screen that the sweater fills a large portion of his field of view; and he also finds the sweater to be green when the image is made small and the observer is at a distance from the screen.

Experiment 8.—The same picture is shown with the red filter placed on the short projector, and with no filter on the long projector, except that the polarizers are turned to give the best color. The sweater is now reddish and the hair greenish, and the lips an intense blue-green, illustrating the phenomenon of hierarchical reversal, usual, to be sure, in any color photography system but particularly important here because it demonstrates that the colors are independent of what the observer expects, and that the system, in spite of use of the new "arbitrary" primaries, behaves like any system of color reproduction.

Experiment 9.—A slide is shown in which there are twelve colored objects photographed against a gray background divided into twelve numbered squares (Fig. 5).

It is pointed out that, except for the orange on square No. 2, each object could appropriately be any color, so that there can be no prior association between the name of the object and the name of the color of the object; thus one would not know that the telephone is red, the paper stapler gray, or the design on the teacup, on square No. 12, dark blue. It is pointed out that this slide had been used to interrogate naive observers, and that, except for the color-blind, marked consistency was found in their reporting of the color names.

Experiment 10.—Using the same standard-objects slide as for Experiment 9, a pair of polarizers is placed in position in front of each projector, the relative brightness is varied from the position in which only the red image shows, to the position where just the white image shows. The audience is asked to indicate the two extremes at which color first appears; that is, with only the red image on the screen, the white is slowly added, and the colors at the point where they first appear are noted. Then, with only the white image on the screen, the red is slowly added, and as the colors appear they are again noted. It is pointed out that when

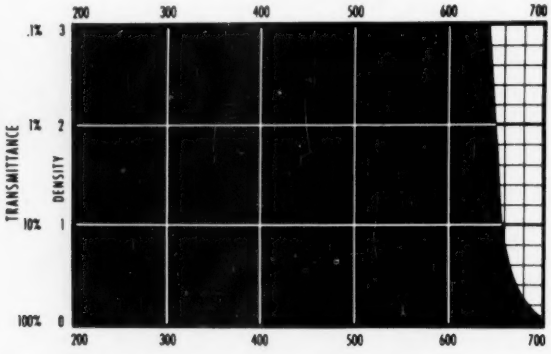

FIG. 8.—Kodak Wratten filter No. 70

color does appear at both extremes it appears at the same time for all the objects on the slide. It is demonstrated that the ratio of stimuli can be varied enormously —with some pairs of stimuli by ratios as high as 100 to 1—without altering the color name given to each one of the objects.

Experiment 11.—In this experiment a duplicate of the short record is superimposed on the short record already in the projector, thus doubling its contrast. A single long record is used, so that it has the original contrast of the previous experiments. Thus for the short record in places where, in the previous experiments, a fifth of the light was transmitted, the slide now transmits a twenty-fifth; in places where a tenth was transmitted, the slide now transmits a hundredth; in places where all the light was transmitted, the slide now transmits all the light. Since no change was made in the long record, the same amount of light is transmitted by it as was transmitted in the earlier experiments. It follows, therefore, that for each object on the slide there is now a new ratio of long to short stimulus, and it would be expected classically that each object would now be a new color. Actually, the objects retain their original color names, demonstrating the stability of the colors

of an image with respect to the contrast of its component images, and illustrating further that color is independent of the quantity of the stimulus defined locally. It is pointed out that we must not let the casualness of the statement that we have doubled contrast in one image without doubling it in the other obscure in our minds the amazing fact that utterly new pairs of ratios give the same colors in the same places.

Experiments 12, 13, 14, and 15 are designed to demonstrate further that color at a point in an image is independent of the wave-length composition of the radiation

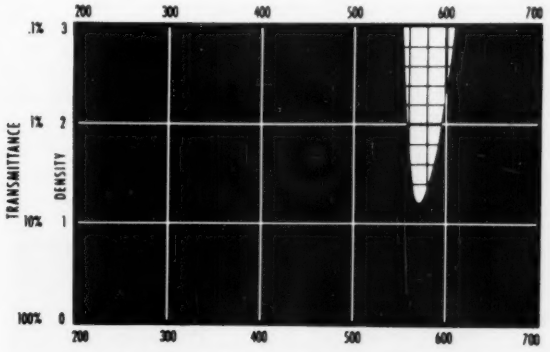

FIG. 9.—Kodak Wratten filter No. 73

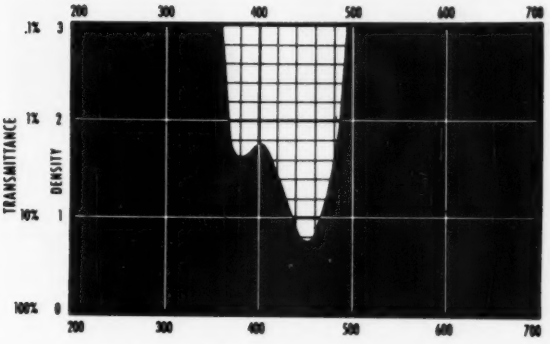

FIG. 10.—Kodak Wratten filter No. 49B

coming from that point. While with changes of filters there is a slight over-all wash of color in extended background areas, objects retain their original color identity. (Many of the audience sitting off the axis of the projector are unable to witness the following four experiments because of the extremely low light level resulting from the dark filters.)

Experiment 12.—A slide showing multicolored groceries on a table top (Fig. 6) is shown with Wratten No. 24 and no filter on the one hand, and on the other hand with Wratten Nos. 24 and 58.

Experiment 13.—A bowl of fruit on a table top (Fig. 7) is shown with Wratten No. 70 (Fig. 8) and Wratten No. 73 (Fig. 9).

Experiment 14.—The slide in Experiment 13 is shown with Wratten No. 73 and Wratten No. 49B (Fig. 10).

Experiment 15.—The same slide is shown with Wratten No. 47B (Fig. 11) and no filter.

Experiment 16.—It is demonstrated to the audience that the filter Wratten No. 73 which passes a narrow band of yellow may be used as a shorter member of a wave-length pair, and as a longer member of a wave-length pair, depending on what is picked to go with it. (*Note:* Experiments 13 and 14.)

ACHROMATIC PHENOMENA

Experiment 17.—Three identical black-and-white transparencies are used, two in superposition on each other in one projector, one in the other projector. These are projected with red and green filters, and the question arises: What colors should one expect? One expects color for two reasons, either classically because of the

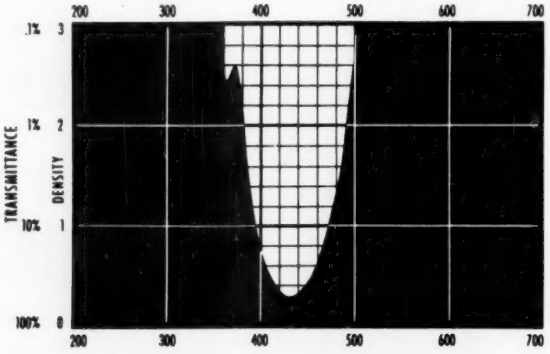

FIG. 11.—Kodak Wratten filter No. 47B

varieties of ratios of red to green or in terms of our new approach because the numerical value of many of the ratios are equal to ratios which existed in the image which showed color in Experiment 9. Actually, one finds that the whole slide is uniform in color and nearly colorless, even though many of the ratios of stimuli for the various objects are the very ratios that existed when the slide appeared in full color.

Experiment 18.—The short record is placed in the short projector and no record is placed in the long projector. One sees the wash of red light over the image from the short projector. The polarizer on the long projector is varied over its range. In spite of the fact that all possible ratios of stimuli have been on the screen, there appears only a uniform wash of the colors in their classical mixtures.

Experiment 19.—Another class of achromatic phenomena is introduced by this experiment in which a photographic step-wedge, running horizontally, is placed in one projector, and a step-wedge, running vertically, is placed in the other projector. Thus, in superposition on the screen, the wedges create 256 possible ratios of the

two stimuli, arranged, however, in a geographically systematic way. The colors mix only in a classical sense, red and white giving a vague pink, and red and green showing some red, some green, and some yellow.

Still another class of nearly colorless situations is exemplified by Experiments 20 and 21.

Experiment 20.—The map slide (Fig. 4) is projected with the long and short record far out of registration. The gamut of colors is strongly restricted. While going beyond the classically expected colors, there is only the smallest suggestion of the new colors which here are few and unsaturated. The images are slowly brought into registration. At the instant of perfect registration, the full gamut of color snaps into appearance.

Experiment 21.—A picture of one scene is placed in the long projector, and a picture of a different scene is placed in the short projector. The limitations on color are similar to those in Experiment 20: there is almost none.

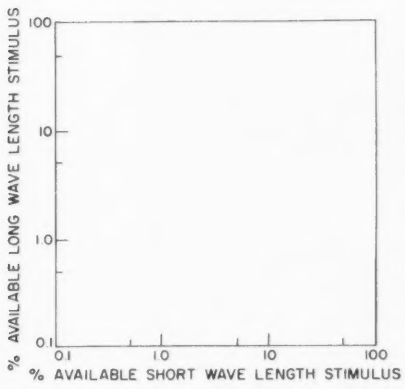

FIG. 12.—Co-ordinate system

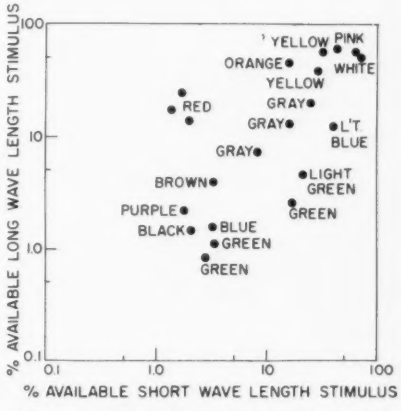

FIG. 13.—Co-ordinate system with loci of colors

A NEW CO-ORDINATE SYSTEM

A new co-ordinate system is shown which describes color in images and which predicts the achromatic situations (Fig. 12). A remarkable thing about the new co-ordinate system is that it is physically dimensionless, involving a ratio of ratios and holding constant an arbitrary pair of wave-lengths. In this co-ordinate system the percentage of available short-wave stimulus is plotted against the percentage of available long-wave stimulus.[3] The first message of the co-ordinate system is that color in an image is indeed independent of the over-all flux in the individual component images and, second, that the color at a point in an image is independent of the wave-lengths of the radiation at that point. This is true of what is seen with a large variety of pairs of wave-lengths and relative brightnesses. Part II of this paper will show that the gamut of colors is restricted when some pairs of wave-lengths are used, and that there are levels of relative energies at which there are significant color shifts. However, even in these situations, the color is not at all the classical function of wave-length and relative energy.

We now study our twenty-one experiments in terms of the co-ordinate system. Slides such as those in Figures 3, 4, 5, 6, 7, and 17 are measured as indicated above, and the results are shown in Figure 13. Here we see that in this new dimensionless co-ordinate system the grays fall on a straight line at an angle of 45° to the axes, warm colors being to one side of the gray and cools to the other.

Doubling the contrast of the image, as in Experiment 11, rotates the gray line, which carries with it the family of warm and cool colors, disposed as they were originally on either side of the gray line; in Figure 14 imaginary lines have been drawn through some of the colored areas, the unbroken lines representing points in an image of normal contrast, the dotted lines indicating the new position of those points when the contrast of one of the images has been doubled. The significant fact is that the colors themselves remain constant for the observer, in spite of this translation in the co-ordinate system, as if the eye were indifferent to the change

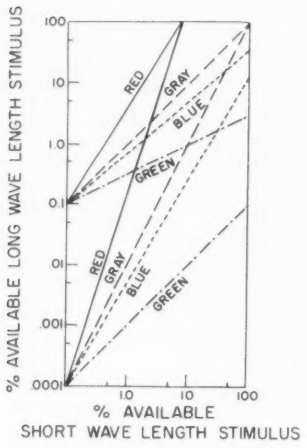

FIG. 14.—Elastic transformation of co-ordinate system effected by doubling contrast in one image.

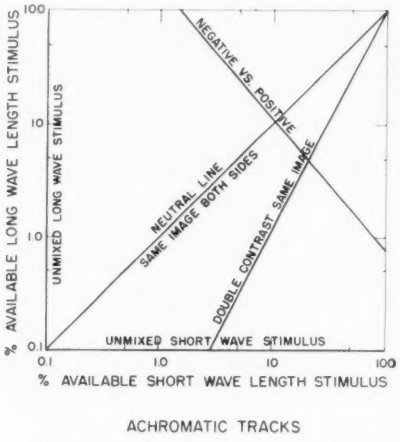

FIG. 15.—Co-ordinate system showing straight lines representing ratios in achromatic images.

in position of the array as a whole as long as the relative positions of the color points in the array remain unchanged.

Achromatic Situations.—When the points in the images in Experiments 17 and 18 are plotted on the diagram they are found to fall on straight lines (Fig. 15), suggesting that this is the criterion for whether or not an image is achromatic, and possibly suggesting further that if the points measured in an image fall on any smooth continuum in the diagram, such an image may also be achromatic. For three identical images in Experiment 17 all the points fall on a line (which is in the same position as the gray line in the double-contrast image of Experiment 11). When a single image in one projector is combined with a wash of light in another, as in Experiment 18, the measured points will fall on the straight line parallel to one of the axes. As the observer's eye moves over the field of the crossed wedges in Experi-

ment 19, in any direction, the points traversed, when measured, will fall on simple smooth curves on the diagram of the co-ordinate system.

If the wedges were continuous instead of being step-wedges, then one might think of the crossed wedges as being the continuum which is the co-ordinate diagram as a whole.

The last class of situations, which are nearly colorless, includes the out-of-register images of Experiment 20 and the two different images of Experiment 21. If one thinks of the fusion of the components of the colored image on the retina as being analogous to the central fusion of the two members of the stereoscopic pair on the two different retinas, then it seems plausible that the sense of color might be greatly

FIG. 16.—Spectrum of natural objects, negative and positive

diminished when the component images on one retina are out of register, just as the three-dimensional sense will be greatly diminished, indeed destroyed, if the members of the stereoscopic pair do not lie in "registry" on the corresponding points of the two retinas. A second way of thinking, which will not be discussed in this paper, has to do with the relationship between the way the eye moves over the geography of the image and the trace of the point on the geometry of our diagram.

Experiment 22.—To test the hypothesis that continua in general, and straight lines in particular, have a special significance in our co-ordinate system, one may ask the following question: If the measured points in the image fall on a straight line at *right* angles to the 45° line in the co-ordinate system, what colors will appear? Now if a negative image is placed in one projector and the positive from that

negative is placed in the other projector (Fig. 16), then the measured points *will* fall on a line at right angles to the 45° line. The red filter is placed on the long projector and no filter is placed on the short projector; the negative and positive images are registered carefully. Is the co-ordinate system a good prophet? All ratios of stimuli are present in these images which are geometrically identical and in register. Will this be an image in good color, or will it be essentially colorless? The answer is that the co-ordinate system *is* a good prophet. There are only reds and whites, and a little pinkness. This is further support for the hypothesis that stimuli whose measurements are disposed on a straight line in our co-ordinate system do *not* "mix" to give color. (The control picture, the two positives taken and

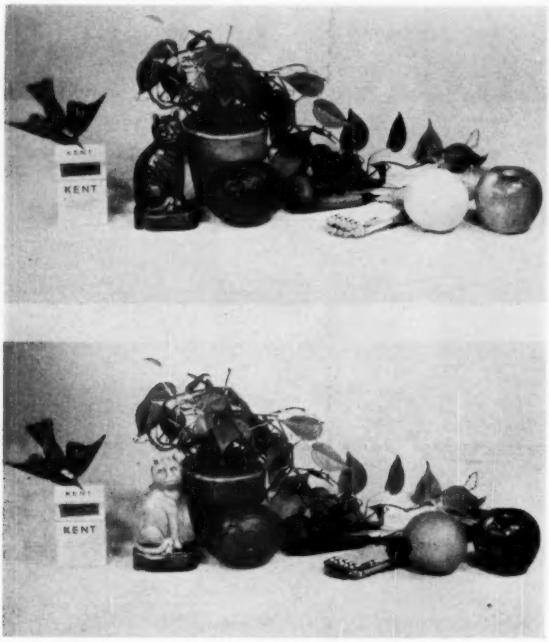

FIG. 17—Spectrum of natural objects, control.

shown in our new "standard" way [Fig. 17], is vividly colored.) This negative-positive pair promises to be one of the great clues to the nature of color vision. It will be discussed in the next part of this paper.

Because of the limitations of flexibility imposed by using filters, a dual monochromator has been built since the time of the symposium. The work with this instrument has supported the experiments shown at the symposium and has enabled us to arrive at a meticulous description of the separation of wave-lengths, as a function of wave-length, that is required to see the gamut of colors for experiments like those described above. This instrument and the detailed results obtained with it will be described in Part II. There will also be described new apparatus demonstrated at the meeting of the Optical Society of America on October 9, 1958, which

enables an audience to see that the same monochromatic stimulus can be either the long or short member of a pair of stimuli. The demonstrations given at the Physics Colloquium, Harvard University, on November 17 and 18, 1958, showing how the eye functions to see full color in a universe involving only a narrow wave-length band ("Johnny's Magic Bookcase") will be described, and the generalization will be drawn that our everyday world is just a special case among these universes.

Part III will continue the discussion of the nature of the field in the image situation and the relationship of the spectrum to color vision.

* This paper will be in several parts, the first of which is this symposium. Part I then contains the experiments up to the time of the symposium, April 28, 1958.

[1] The work in this paper should not be confused with experiments such as those of Hess, which describe how a stimulus of a particular wave-length looks differently colored when viewed after the eye has been exposed to another wave-length for a significant period of time (*Arch. Ophthalmol.*, **36, I** [1890]). Nor should it be confused with those theories which explain the phenomenon which Hess measured, and other allied phenomena, as the result of fatigue of the eye to various states of chromatic adaptation.

The colors described in our experiments appear immediately, and do not alter appreciably with time, and we do not seek to explain them as effects of adaptation. To the best of our knowledge the gamut of color is much larger than that noticed by Hess or predicted by any theory concerning chromatic adaptation; our newly discovered achromatic situations are not predicted by any previous theories; nor is there any basis in those theories for our dimensionless co-ordinate system (see below). Furthermore it seems that none of the previous work in any kind of color matching or system of complementary colored shadows (Goethe) can be the basis for the observed fact that the colors in our images, arising from unorthodox stimuli, correspond in hierarchical order to the colors that the observer would have seen had he been at the original site.

[2] Earlier studies in two-color projection were done by Fox and Hickey who found that moving pictures which showed some color could be taken by alternately photographing through no filter and through a filter which transmitted red light, and by projecting correspondingly (British Patent No. 636, issued July, 1914, to William Francis Fox and Wm. H. Hickey, *Improvements in Kinematographic Apparatus*). Some years later Bernardi used red and green filters for taking movies and red and white for projecting (Adrian Cornwell-Clyne, *Colour Cinematography* [London, 1951], p. 261).

[3] This means that one measures the amount of light on the screen for the long wave-length image in the brightest place and takes that as 100; then one measures the amount of light for that long record alone at another place in the image and makes the percentage. Then that percentage is plotted against the percentage of available light for the short record alone at the same point on the screen. The ratio of these two percentages describes the color at that point.

INFORMATION TO CONTRIBUTORS

THE PROCEEDINGS is the official organ of the NATIONAL ACADEMY OF SCIENCES and of the NATIONAL RESEARCH COUNCIL for the publication of brief accounts of important current researches of members of the ACADEMY and of the COUNCIL and of other American investigators. THE PROCEEDINGS will aim especially to secure prompt publication of original announcements of discoveries and wide circulation of the results of American research among investigators in other countries and in all branches of science.

ARTICLES should be brief. The viewpoint should be comprehensive in giving the relation of the paper to previous publications of the author or of other and in exhibiting, where practicable, the significance of the work for other branches of science. Elaborate technical details of the work and long tables of data should be avoided, but authors should be precise in making clear the new results and should give some record of the methods and data upon which they are based.

MANUSCRIPTS should be prepared with a current number of THE PROCEEDINGS as a model in matters of form, and should be typewritten in duplicate with double spacing, the author retaining one copy. Illustrations should be confined to text figures of simple character, though more elaborate illustration may be allowed in special instances to authors willing to pay for their preparation and insertion. Particular attention should be given to arranging tabular matter in a simple and concise manner.

MANUSCRIPTS will be accepted only from members of the Academy who will assume responsibility for the propriety and scientific standards of the paper and for any printing costs in excess of those allowed them by the Academy. Manuscripts should be addressed to THE PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES, 2101 Constitution Ave., Washington 25, D. C.

GALLEY PROOF will be sent, with the understanding that charges for author's corrections shall be billed to him. Author should therefore make final revisions on the typewritten manuscripts. Original artwork or figures cannot be returned unless a request is made when the manuscript is submitted. Page proofs will not be sent.

REPRINTS should be ordered when proof is returned, on the reprint order form which will be inclosed with the proof. They will be furnished at cost, approximately as follows:

SCHEDULE OF REPRINT RATES (POSTAGE NOT INCLUDED)

Quantity	up to 4 pp.	up to 8 pp.	up to 12 pp.	Standard Covers
50	$ 7.95	$15.10	$21.50	$ 6.95
100	9.05	17.45	25.85	8.50
150	10.15	19.65	29.20	10.00
add'l 100's	2.25	4.45	6.65	3.05

CONTENTS

Page

BIOCHEMISTRY.—A MITOCHONDRIAL FACTOR PRODUCING "LOOSE-COUPLING" OF RESPIRATION . . . . . *By LeMar F. Remmert and Albert L. Lehninger* 1

BIOCHEMISTRY.—GLYCOGEN SYNTHESIS FROM GLUCOSE, GLUCOSE-6-PHOSPHATE, AND URIDINE DIPHOSPHATE GLUCOSE IN MUSCLE PREPARATIONS . . . . . *By Phillips W. Robbins, Robert R. Traut, and Fritz Lipmann* 6

CHEMISTRY.—DISSOCIATION OF HUMAN CARBONMONOXYHEMOGLOBIN AT HIGH pH . . . . . *By Ulrich Hasserodt and Jerome Vinograd* 12

GENETICS.—RADIATION INDUCED REVERSE MUTATIONS IN *Drosophila melanogaster* . . . . . *By M. M. Green* 16

GENETICS.—DEPENDENCE OF MUTATION FREQUENCY ON RADIATION DOSE RATE IN FEMALE MICE . . . . . *By W. L. Russell, Liane Brauch Russell, and Mary B. Cupp* 18

GENETICS.—MODIFICATION OF IRRADIATION EFFECTS IN THE PIGEON, *Columba livia*. I. EFFECTS OF HOMOLOGOUS AND HETEROLOGOUS BONE MARROW TRANSPLANTS . . . . . *By Donald H. Shaw and Halvor Vermund* 23

GENETICS.—ON THE NATURE OF THE AGING PROCESS . . . . . *By Leo Szilard* 30

MATHEMATICS.—ON IDENTICAL VANISHING OF HOLOMORPHIC FUNCTIONS IN SEVERAL COMPLEX VARIABLES . . . . . *By S. Bochner* 46

MATHEMATICS.—LINEAR AND ALGEBRAIC DEPENDENCE OF FUNCTIONS ON COMPACT COMPLEX SPACES WITH SINGULARITIES . . . . . *By S. Bochner* 47

MATHEMATICS.—LINE-ELEMENT FIELDS ON THE TORUS . . . . *By Bruce L. Reinhart* 49

MATHEMATICS.—ON A PROBLEM OF MAZURKIEWICZ CONCERNING THE BOUNDARY OF A COVERING SURFACE . . . . . *By Peter Seibert* 50

PATHOLOGY.—THE EFFECTS OF STRONTIUM-90 ON MICE . . . . . *By Barclay Kamb and Linus Pauling* 54

PHYSICS.—GRAVITATIONAL PROPERTIES OF ANTIMATTER . . . . . *By L. I. Schiff* 69

PHYSIOLOGY.—SPECIFIC INTRACELLULAR BINDING OF RUBIDIUM BY RAT DIAPHRAGM MUSCLE . . . . *By P. Menozzi, D. Norman, A. Polleri, G. Lester, and O. Hechter* 80

PHYSIOLOGY.—SYMPOSIUM ON NEW DEVELOPMENTS IN THE STUDY OF COLOR VISION . . . . . *H. K. Hartline*

ABOUT THEORIES OF COLOR VISION . . . . . *By Yves LeGrand* 89

STUDIES OF COLOR BLINDNESS: A UNILATERALLY DICHROMATIC SUBJECT . . . . . *By C. H. Graham and Yun Hsia* 96

COLOR VISION: THE APPROACH THROUGH INCREMENT-THRESHOLD SENSITIVITY . . . . . *By W. S. Stiles* 100

VISUAL PIGMENTS IN THE INTACT HUMAN EYE . . . . . *By W. A. H. Rushton* 114

COLOR VISION IN THE NATURAL IMAGE . . . . . *By E. H. Land* 115